

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 700.638 Seconds
(without alignments)
7148.662 Million cell updates/sec

Title: US-09-855-340A-1
Perfect score: 1179
Sequence: 1 ggtgagatcgagaagaacgg.....tggggcgcgcagcatga 1179

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_25Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1179	100.0	1179	6	AAD25931
2	1179	100.0	4388	5	AAS08694
3	61.6	5.2	2404	2	AA337131
4	60.4	5.1	2000	7	ADA71938
5	56.6	4.8	1161	6	AD21926
6	56.6	4.8	1188	6	AD21912
7	55.8	4.7	27541	4	AD17185
8	55.8	4.7	125401	4	AD17186
9	55.2	4.7	114955	2	AA53491
10	55	4.7	114955	2	AA53491
11	53.2	4.5	2000	7	ADA71938
12	53	4.5	504	7	AD71938
13	53	4.5	82746	7	AD61190
14	52.8	4.5	24379	2	AA93095
15	52.8	4.5	24379	2	AA93095
16	52	4.4	88421	6	AA25925
17	51.4	4.4	516	2	AA47574
18	51.4	4.4	516	6	AA47574
19	51.4	4.4	516	6	AA47574
20	51.4	4.4	516	6	AA47574
21	51.4	4.4	516	6	AA47574
22	51.4	4.4	516	6	AA47574
23	51.4	4.4	582	6	AA47574

24	51.4	4.4	582	8	ADB78877
25	51.4	4.4	7065	6	AAF88577
26	51.4	4.4	7065	8	ADB78878
27	50.8	4.3	2136	7	ACA37643
28	50.6	4.3	1677	7	AA161210
29	50.6	4.3	82746	7	AA161224
30	50.4	4.3	11238	7	AA161224
31	50.4	4.3	47981	4	AA55817
32	50.4	4.3	60196	7	AA55817
33	50	4.2	1291	4	AA55810
34	50	4.2	3498	4	AA55810
35	50	4.2	3790	4	AA55810
36	49.8	4.2	4770	7	AA55810
37	49.8	4.2	109519	5	AA55810
38	49.6	4.2	9222	7	AA55810
39	49.6	4.2	135638	7	AA55810
40	49	4.2	3197	6	AA55810
41	49	4.2	14055	7	AA55810
42	48.6	4.1	1137	7	AA55810
43	48.6	4.1	1189	7	AA55810
44	48.6	4.1	1290	7	AA55810
45	48.6	4.1	53789	2	AA55810

ALIGNMENTS

RESULT 1
AAD25931
ID AAD25931 standard; DNA; 1179 BP.
XX AC AAD25931;
XX AC AAD25931;
DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 integrase DNA.
XX DE Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX DE site-specific integration; hybrid antibiotic; metabolic product;
XX DE secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX FN WO200187936-A2.
XX PD 22-NOV-2001.
XX PF 15-MAY-2001; 2001WO-US015760.
XX PR 17-MAY-2000; 2000US-0204670P.
XX (SCHE) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX Novel polynucleotides encoding integrase, excisionase and an integrase attachment site isolated from a lysogenic phage pMLP1, useful for transforming an actinomycete.
XX Claim 5; Page 32; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase (int) and excisionase (xis) and an integrase attachment site (attP) which are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from Micromonospora carbonacea var. africana. Polynucleotides of the invention are useful for transforming an actinomycete with a vector. They are also useful for creating vectors for site-specific integration into host chromosomes. The integrating vectors are used to express actinomycete genes, manipulate secondary metabolic pathways and create new metabolic products such as hybrid antibiotics. The present sequence is pMLP1 integrase DNA from Micromonospora carbonacea

```
XX SQ Sequence 1179 BP; 214 A; 370 C; 410 G; 185 T; 0 U; 0 Other;
Query Match 100.0%; Score 1179; DB 6; Length 1179;
Best Local Similarity 100.0%; Pred. No. 6.2e-207;
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGATCGAGAGAACGGCCGCTTACCGATTTCGGGACCTCGTTCCGGTAAAG 60
DB 1 GTGTGATCGAGAGAACGGCCGCTTACCGATTTCGGGACCTCGTTCCGGTAAAG 60
QY 61 GTACCATTCAGACCGGTTATCCAGCAAGACCAAGCCCAAGATGCGATGTC 120
DB 61 GTACCATTCAGACCGGTTATCCAGCAAGACCAAGCCCAAGATGCGATGTC 120
QY 121 CGTGGGAGCAGTTGTCAGGGCAACGCGCTCATGCGCGCGCGGCTCAGATTACCTCGCC 180
DB 121 CGTGGGAGCAGTTGTCAGGGCAACGCGCTCATGCGCGCGCGGCTCAGATTACCTCGCC 180
QY 181 GATTTCGTGGGGAGTGTGTGCGAGCTACGAAAGACGCTGAAACCGACCGCGTGAAC 240
DB 181 GATTTCGTGGGGAGTGTGTGCGAGCTACGAAAGACGCTGAAACCGACCGCGTGAAC 240
QY 241 TCGGAGGCAACCGGATCCGCAACACCTCTGCGCCATCTCGGCGCATCTCACCCCTGAC 300
DB 241 TCGGAGGCAACCGGATCCGCAACACCTCTGCGCCATCTCGGCGCATCTCACCCCTGAC 300
QY 301 GAGTGTGACGGGAGTGTGACCAAGTGTGCTCAAGCTGAGCGCGGCTGCGCCG 360
DB 301 GAGTGTGACGGGAGTGTGACCAAGTGTGCTCAAGCTGAGCGCGGCTGCGCCG 360
QY 361 TGGCGGAGTCCAGCGGGGTCGTCGAAGCGCTGGCAGCAAGACGATCAGCAACTGC 420
DB 361 TGGCGGAGTCCAGCGGGGTCGTCGAAGCGCTGGCAGCAAGACGATCAGCAACTGC 420
QY 421 CACGCGCTGTGCAACAGATCTGCGCGCGCGGATCGCGCGAAGCGGATCAGGCTCAAC 480
DB 421 CACGCGCTGTGCAACAGATCTGCGCGCGCGGATCGCGCGAAGCGGATCAGGCTCAAC 480
QY 481 CCGTGTCTTCGACGATGCTGCGCGCGCGGAGCGCAAGAGATGAATTCCTGAGCGAC 540
DB 481 CCGTGTCTTCGACGATGCTGCGCGCGCGGAGCGCAAGAGATGAATTCCTGAGCGAC 540
QY 541 CCGGAGATCGGTGCGCTTATCAGCGCGCTTCCGCGCGCTTCCGCGCGCTGCGTCAAGTGTG 600
DB 541 CCGGAGATCGGTGCGCTTATCAGCGCGCTTCCGCGCGCTTCCGCGCGCTGCGTCAAGTGTG 600
QY 601 CTGTGTGCGCGCGGCTGAGTGGGTGAGCGGATCGGCGCGCGCGGCTGCGAC 660
DB 601 CTGTGTGCGCGCGGCTGAGTGGGTGAGCGGATCGGCGCGCGCGGCTGCGAC 660
QY 661 CTGTGTGCGCGCGGCTGAGTGGGTGAGCGGATCGGCGCGCGCGGCTGCGAC 720
DB 661 CTGTGTGCGCGCGGCTGAGTGGGTGAGCGGATCGGCGCGCGCGGCTGCGAC 720
QY 721 GGAGAGTCTGCTTCCAGTCCGCAAGACCGCAAGCGCGCGCGCGCTCAGTTTCAAC 780
DB 721 GGAGAGTCTGCTTCCAGTCCGCAAGACCGCAAGCGCGCGCGCGCTCAGTTTCAAC 780
QY 781 ACAGAGTCTGCTTACTGCTTACGCCACTCATCGCGGAAAGAAAGTGAAGGTCGTG 840
DB 781 ACAGAGTCTGCTTACTGCTTACGCCACTCATCGCGGAAAGAAAGTGAAGGTCGTG 840
QY 841 TTACCGCGCGCGGAGGATGTAAGACCGCAATTTCCGCGCGGATCTGGGTCAAG 900
DB 841 TTACCGCGCGCGGAGGATGTAAGACCGCAATTTCCGCGCGGATCTGGGTCAAG 900
QY 901 CGGTGCGAGGAGCGCGGCTTCCGGCTTACGATTCAGATTCGCGGCACTCAGCGG 960
DB 901 CGGTGCGAGGAGCGCGGCTTCCGGCTTACGATTCAGATTCGCGGCACTCAGCGG 960
QY 961 GCGATCTGATTTCTGCGCGGCGTCCGCTGTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020
DB 961 GCGATCTGATTTCTGCGCGGCGTCCGCTGTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020

Db 961 GCGATCTGATTTCTGCGCGGCGTCCGCTGTGCGGATCTCCCGCGCGCTCGGTCACTCG 1020
QY 1021 TCGATCCGCTCAGCGATCTGTTACGGGCACTTGGGACCTTGGAGAGTGGACAGGGGATC 1080
DB 1021 TCGATCCGCTCAGCGATCTGTTACGGGCACTTGGGACCTTGGAGAGTGGACAGGGGATC 1080
QY 1081 CTGCGGCGGATCGAGAGGCGATGGCGCGCTCGGCGCTGAGGACCTGGAGCGGAACTC 1140
DB 1081 CTGCGGCGGATCGAGAGGCGATGGCGCGCTCGGCGCTGAGGACCTGGAGCGGAACTC 1140
QY 1141 GACAGAGCTGACGAGCGTGTGGCGGACGAGCATGA 1179
DB 1141 GACAGAGCTGACGAGCGTGTGGCGGACGAGCATGA 1179

RESULT 2
AAS08694
ID AAS08694 standard; DNA; 4388 BP.
XX AC AAS08694;
XX DT 11-SEP-2003 (revised)
XX DT 26-SEP-2001 (first entry)
XX XX Micromonospora DNA encoding integrase enzymes.
XX KW Evernimycin; antibiotic; bottle-neck gene; orthomycin; fermentation;
XX KW integrase; ds.
XX OS Micromonospora sp. ATCC 39149.
XX PH Key Location/Qualifiers
XX CDS 963..1387
XX FT /*tag= a
XX FT /product= "Integrase #1"
XX FT 1394..2572
XX FT /*tag= b
XX FT /product= "Integrase #2"
XX FT misc_feature 2570..2799
XX FT /*tag= c
XX FT /note= "AttB/AttP region of integrase action"
XX FT 27114..2715
XX FT /*tag= d
XX FT misc_recomb
XX FT /label= Insertion_juncture
XX FT /note= "Site of integrase activity"
XX PN WO200151639-A2.
XX PD 19-JUL-2001.
XX PF 12-JAN-2001; 2001WO-US001187.
XX PR 12-JAN-2000; 2000US-0175751P.
XX PA (SCHE ) SCHERING CORP.
XX PI Hosted TJ, Horan AC, Wang TX;
XX DR WPI; 2001-442147/47.
XX DR P-PSDE; AAU04900, AAU04912.
XX PT New nucleic acid molecules encoding evernimycin pathway gene products,
XX PT useful for improving yields of evernimycin, to produce new
XX PT evernimycin and as probes to identify homologous sequences.
XX PS Claim 26; Fig 7; 109pp; English.
XX CC The sequence encodes 2 integrases which permit site specific integration
XX CC of a vector into an actinomycete, especially a Micromonospora, genome.
XX CC The invention relates to nucleic acids and vectors comprising a M.
XX CC carbonacea evernimycin biosynthetic pathway resistance gene product
XX CC useful for selecting for a transfected or transformed host cell. An
XX CC integrative version of the vector is useful for introducing a
```

CC everninomicin pathway gene (a bottle-neck gene) into an actinomycete of
 CC the genus Micromonospora. The DNA encoding the biosynthetic proteins is
 CC useful for synthesizing novel everninomicin-related compounds, arising
 CC from modifications of the DNA sequence designed to change glycosyl and
 CC modified orsellinic acid groups contained in everninomicin, for
 CC expressing functional or mutant everninomicin biosynthetic enzyme for
 CC evaluation, diagnosis and preferably biosynthesis of everninomicin or
 CC other secondary metabolic products, improving the yield of everninomicins
 CC and to produce novel everninomicins, and also as a hybridisation probe to
 CC identify homologous sequences. The encoded polypeptides are useful for
 CC combinatorial biosynthesis to generate libraries of orthomycins, e.g.
 CC everninomicin analogues/homologues and drug discovery. The DNA encoding
 CC the integrase allows for increasing a given gene dosage. The integrative
 CC vector can be used to permanently integrate copies of a heterologous gene
 CC of choice into chromosomes of different hosts and to integrate genes
 CC which increase the yield of known products or to generate novel products
 CC such as hybrid antibiotics or other novel secondary metabolites. The
 CC vector can also be used to integrate antibiotic resistance genes in order
 CC to carry out bioconversions with compounds to which the strain is
 CC normally sensitive and is thus useful in fermentation processes involving
 CC e.g. Streptomyces antibiotics. (Updated on 11-SEP-2003 to standardise OS
 CC field)
 XX
 SQ Sequence 4388 BP; 714 A; 1468 C; 1506 G; 694 T; 0 U; 6 Other;

Query Match 100.0%; Score 1179; DB 5; Length 4388;
 Best Local Similarity 100.0%; Pred. No. 6.1e-207;
 Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGATCGAGAGAAACGGGCGCTCTACCGCATTCGGGACCTCGTTCCGGGTAAAAAG 60
 DB 1394 GTGTGATCGAGAGAAACGGGCGCTCTACCGCATTCGGGACCTCGTTCCGGGTAAAAAG 1453
 QY 61 GTCCACATTCAGACCGGTATCCGACGAAGACAGCGCCCAAGATGCGATGTCAGTTTC 120
 DB 1454 GTCCACATTCAGACCGGTATCCGACGAAGACAGCGCCCAAGATGCGATGTCAGTTTC 1513
 QY 121 CGTGGGAGCTTCAGGGCAACCGCTCATCGCGCGGGCGGTTCAGATTACCTCGCC 180
 DB 1514 CGTGGGAGCTTCAGGGCAACCGCTCATCGCGCGGGCGGTTCAGATTACCTCGCC 1573
 QY 181 GATTTCGTGGGGAGTGTGGCGGAGTACGAAAGAGCGCTGAACACCGCCGCTGAAC 240
 DB 1574 GATTTCGTGGGGAGTGTGGCGGAGTACGAAAGAGCGCTGAACACCGCCGCTGAAC 1633
 QY 241 TCGGAGGCAACCGGATCGGAACCACTCTCTGCCATCTCGGCCATCTCACCTTTGAC 300
 DB 1634 TCGGAGGCAACCGGATCGGAACCACTCTCTGCCATCTCGGCCATCTCACCTTTGAC 1693
 QY 301 GAGCTGGACGGGACGTCACCCAGCAGTGGGTCAACGACCTGGAGCGGGCGGTGGCCCG 360
 DB 1694 GAGCTGGACGGGACGTCACCCAGCAGTGGGTCAACGACCTGGAGCGGGCGGTGGCCCG 1753
 QY 361 TGGCGGAGTCCACCGGGGTCTGCGGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 420
 DB 1754 TGGCGGAGTCCACCGGGGTCTGCGGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 1813
 QY 421 CACGSCCTCTGCACACCATCTCGGCGCGGGATTCGCGGCAACCGATCAGGCTCAAC 480
 DB 1814 CACGSCCTCTGCACACCATCTCGGCGCGGGATTCGCGGCAACCGATCAGGCTCAAC 1873
 QY 481 CCGTCTCTTCGACGATCTCCCGCGCGGAGCGCGAAGAGATCAAGTTCTCTGAGCGAC 540
 DB 1874 CCGTCTCTTCGACGATCTCCCGCGCGGAGCGCGAAGAGATCAAGTTCTCTGAGCGAC 1933
 QY 541 CCGGAGATCGGTGGCTTATCAGCGGCTTCGCGCGGACATGCGGACCGCTCGTCACTG 600
 DB 1934 CCGGAGATCGGTGGCTTATCAGCGGCTTCGCGCGGACATGCGGACCGCTCGTCACTG 1993
 QY 601 CTGCTGGCGACCGGTCTGAGTGGGGTTCAGCGCATCGGCTCGCGCGGCGGGTTCGAC 660
 DB 1994 CTGCTGGCGACCGGTCTGAGTGGGGTTCAGCGCATCGGCTCGCGCGGCGGGTTCGAC 2053

QY 661 CTGCTCGCGCGCGGCGCCCGCTGACCGTCTGAGCAGCTCCAGAGCTGGCCAGCAAG 720
 DB 2054 CTGCTCGCGCGCGGCGCCCGCTGACCGTCTGAGCAGCTCCAGAGCTGGCCAGCAAG 2113
 QY 721 CGAGAGCTCGTCTTCCAGTCCGCCAAGACCGCAAGGCGCGGCGCAGGTCAGTTTCACC 780
 DB 2114 CGAGAGCTCGTCTTCCAGTCCGCCAAGACCGCAAGGCGCGGCGCAGGTCAGTTTCACC 2173
 QY 781 ACGAAGTCTGCTTACTGCTTACGCCACTCATCCCGGAAAGTGCAGAGTCTGTG 840
 DB 2174 ACGAAGTCTGCTTACTGCTTACGCCACTCATCCCGGAAAGTGCAGAGTCTGTG 2233
 QY 841 TTCACCGCGCGCAAGCGGATGTAAGGACGCGCAATTTCGGCGGATCTGGTCAAG 900
 DB 2234 TTCACCGCGCGCAAGCGGATGTAAGGACGCGCAATTTCGGCGGATCTGGTCAAG 2293
 QY 901 GCGTGCAGGAAGCGGGCTTCGGGGTTACGCAATTCACGATCTGCGGCGACACTCAGCGG 960
 DB 2294 GCGTGCAGGAAGCGGGCTTCGGGGTTACGCAATTCACGATCTGCGGCGACACTCAGCGG 2353
 QY 961 GCGATCTGATTTCTGCGGGCGTCCGCTGTCGGCGATCTCCCGCGCTCGTCACTCG 1020
 DB 2354 GCGATCTGATTTCTGCGGGCGTCCGCTGTCGGCGATCTCCCGCGCTCGTCACTCG 2413
 QY 1021 TCGATCGCGTTCAGGATCTGCTTACGGCACCTGCGTGAGGAGTCCGACGAGGATC 1080
 DB 2414 TCGATCGCGTTCAGGATCTGCTTACGGCACCTGCGTGAGGAGTCCGACGAGGATC 2473
 QY 1081 CTGCGCGGATCGAGGAGGATGCGCGGCTCGGGCTCGGCTGAGGACCTGAGGCGGAATC 1140
 DB 2474 CTGCGCGGATCGAGGAGGATGCGCGGCTCGGGCTCGGCTGAGGACCTGAGGCGGAATC 2533
 QY 1141 GACGAGGAGTGCAGGACGCTGTTGGCGACGCGACATGA 1179
 DB 2534 GACGAGGAGTGCAGGACGCTGTTGGCGACGCGACATGA 2572
 RESULT 3
 AAX37131
 ID AAX37131 standard; DNA; 2404 BP.
 XX AC AAX37131;
 XX DT 05-JUL-1999 (first entry)
 XX DE Mycobacteriophage M56 DNA attP-int region.
 XX DE DNA integration; Mycobacterium; bacteriophage; phage attachment site;
 XX DE attP; promoter; integrase; recombinant; transformation efficiency;
 XX OS Mycobacterium sp.
 XX PN WO9907861-A1.
 XX PD 18-FEB-1999.
 XX PF 06-AUG-1997; 97WO-PT000005.
 XX PR 06-AUG-1997; 97WO-PT000005.
 XX (MEDI-) LAB MEDINFAR-PROD FARMACEUTIC LDIA.
 XX PI Frazao Moniz Pereira JA, Freitas Vieira A;
 XX PI Ribeiro Dos Santos Anes EM, Da Costa Garcia MA, Da Silva Alves PJ;
 XX DR WPI; 1999-180493/15.
 XX DR P-PSDB; AAV09007.
 XX PT A new system for integrating DNA into mycobacterium species - allows the
 XX PT stable construction of a vaccine vehicle for long-term antigen gene
 XX PT expression.

Db 473 KAAYSGMMYWRKYKSKWMRMTXYMWSMYKKRSNKYGAAGCYGCKWTCYSGYMKW 414
 Qy 441 CTGCGGCGGCGGATCGCGGGAACGATCAGGCTCAACCGGTGCTCTCGAGATGCT 500
 Db 413 YTYGSKYKRCYKMYKMYKGMWYMYYSAYSSMMTWYYYAKYKMYKRGTMWY 354
 Qy 501 GCCCGGCGGAGCGGAAAGAGATGAAGTCTTCGAGCGACCCGAGATCGGTCTTAT 560
 Db 353 GKSYKKYCTWVCYMKCMRCYWRKQMKKTKYKRCYOWRYATCYWCCYKRGWYSR 294
 Qy 561 CAGGCGCTTCGCGGCACTGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 620
 Db 293 SMRTAGKWRWRSWRCRSYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMYKMY 234
 Qy 621 GTGGGTGAGGCGATCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
 Db 233 STSRRAKQWACRMYKACRYSRTSYCGSYGSSKWKYMSKSMRWTCSWCSCTY 174
 Qy 681 GCTGACGCTGCTGAGCAGCTCCAGGAGCTGCGCAGCAGGAGAGAGCTGCTTCCAGTC 740
 Db 173 CYTGACWCSMSMYTNGSCGYTRGWKWSKYSMCKKYCSCCTKYCSYTGVRKYKY 114
 Qy 741 GCCGAAGACCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
 Db 113 KYSYKCYCYCYWYMSYMYKMKMCRSCSWSWCSWCSWCSWCSWCSWCSWCSW 54
 Qy 801 TAGGCCAC 808
 Db 53 GSSGMYRM 46

RESULT 5

AAD21926
 ID AAD21926 standard; DNA; 1161 BP.

XX AC AAD21926;
 XX DT 12-FEB-2002 (first entry)

XX DE M. rosaria pMR2 plasmid site-specific integrase gene, int.
 XX KW Plasmid pMR2; vector construction; secondary metabolic pathway;
 XX KW antibiotic; int gene; site-specific integrase; ds.
 XX OS Micromonospora rosaria.

XX PN WO200175116-A2.
 XX PD 11-OCT-2001.

XX PF 29-MAR-2001; 2001WO-US010321.
 XX PR 04-APR-2000; 2000US-0194461P.

XX PA (SCHE) SCHERING CORP.
 XX PI Hosted TJ, Horan AC;

XX DR WPI; 2002-010791/01.
 XX XX

XX PT New genes isolated from Micromonospora rosaria plasmid pMR2, useful in
 XX PT constructing vectors for studying and expressing genes, or in
 XX PT manipulating secondary metabolic pathways in actinomycetes.

XX PS Claim 9; Page 33-34; 34pp; English.
 XX CC

CC The invention relates to new genes isolated from Micromonospora rosaria
 CC plasmid pMR2, and proteins encoded by such genes. The isolated genes of
 CC the invention are useful in the construction of vectors, which can be
 CC used in the study and expression of genes, in manipulating secondary
 CC metabolic pathways in actinomycetes and in creating new metabolic
 CC products such as hybrid antibiotics. The isolated genes can also be used

CC to create replicating, Escherichia coli-actinomycete shuttle, integrating
 CC and intermycelial and intramycelial conjugation vectors for use in
 CC actinomycetes. The present sequence is M. rosaria pMR2 plasmid site
 CC specific integrase gene, int. This sequence is involved in pMR2 plasmid
 CC integration
 XX

SQ Sequence 1161 BP; 188 A; 398 C; 387 G; 188 T; 0 U; 0 Other;

Query Match 4.8%; Score 56.6; DB 6; Length 1161;
 Best Local Similarity 46.2%; Pred. No. 0.19;
 Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

Qy 545 AGATCGGTCGGTATATACGGCGCTTCGCGCGCACTGGCGACCGCTCGTCTGCTGG 604
 Db 557 AGTCTGTCGCTTGTGAGGCCATCAACCGCGCTACCGACGCGCTCTGATCGCGG 616
 Qy 605 TGGGACCGGTCTAGGTGGGTGAGCGATCGCGCGCGCGCGCGCGCGCTGCACCTGC 664
 Db 617 CTGTGCGGCTACGCGCGGAGAGATCGCAGACCTCCGGTGCAGACCTGGACCTG- 675
 Qy 665 TCGCGCGCGCGCGCTGACCTGCTGAGCAGCTTCAGGAGCTGCCAGCAGCGGAG 724
 Db 676 --ACCGAACACCATCACCTCCGCAAGCCAGGCTCGAACCGCTGCACGACAGGGGA 733
 Qy 725 AGCTCGTCTTCCAGTTCGCGGAAGACCGGAAGCGCGCGCGCGCGCGCGCTCAGTTTCA 784
 Db 734 AGGCGTTTGACAAGATCCCAAGTCCGAGCGCGGCAAGCAACCATGCCATCCCTCCC 793
 Qy 785 AAGTCGCTTACTCTTACGCCACTATTCGCGCGGAAAGAAAGTACGAGTCTGTTC 844
 Db 794 ACCTCGTACCCGCTGATCCGCTGCACCTCGACGAGTTCCGCGGAAGGATCGCTGTC 853
 Qy 845 CCGCGCGGAAAGCGGATGTTAGGACGCGCAATTTCCGCGGATCTGGTCAAGCGT 904
 Db 854 TCAGCCGCGAGATCACCTTCGCGGCGACACCTGTACGAGCGCTTCGTAGCGCTC 913
 Qy 905 GCGAGGAGCGCGCTTCGCGGTTACGCAATTCAGCATCTCGGCACTCACTCAGCGCGA 964
 Db 914 GGGGAAAGTTCGGAATCGACACCTTCCAGCACTTCGCGCACACCGGTTCAGACCC 973
 Qy 965 TCCTGATTTTCGCGGCGTCCGCTGTCGCGGATCTCCCGCGCTCGTCTGCTCGA 1024
 Db 974 TCGCGCGCAGACCGGCGGACCTTGGCCGACCTGATGAAGCGGCTCGGCACTCGTCCA 1033
 Qy 1025 TCGCGTTCAGGATCTGCTGACGCGCACCTGCTGAGGAGTTCGACGAGGATCTCTCG 1084
 Db 1034 TGGCTG--CGGCTCGCGGTACTCCACGCGCTTGAAGTCTGACGCGGAGATGCCA 1090
 Qy 1085 CGGCGATCGAGGAGGATGCGCGGCG 1111
 Db 1091 AAGCCCTTTCGAGCTGGCGGCGCAG 1117

RESULT 6

AAD21912
 ID AAD21912 standard; DNA; 11188 BP.

XX AC AAD21912;
 XX DT 12-FEB-2002 (first entry)

XX DE Micromonospora rosaria pMR2 plasmid DNA.
 XX KW Plasmid pMR2; vector construction; secondary metabolic pathway;
 XX KW antibiotic; ds.

XX OS Micromonospora rosaria.
 XX CC

XX FT Key Location/Qualifiers
 XX FT misc_feature 59..102
 XX FT /tag= a
 XX FT /note= "pMR2 plasmid attachment site encoding gene, att
 XX FT p"

```

FT misc_feature 269..417
FT /tag= b
FT /note= "DNA replication origin encoding gene"
FT misc_feature 2462..3286
FT /tag= c
FT /note= "Regulatory protein encoding gene, korR"
FT misc_feature 3532..3867
FT /tag= d
FT /note= "Replication regulatory protein encoding gene,
FT ORF11"
FT misc_feature 4408..6072
FT /tag= e
FT /note= "Intramycelial transfer gene, traB"
FT RBS 6055..6059
FT /tag= f
FT /standard_name= "Shine-Dalgarno sequence"
FT misc_feature 5069..5374
FT /tag= g
FT /note= "Intramycelial transfer gene, ORF101"
FT RBS 6391..6394
FT /tag= h
FT /standard_name= "Shine-Dalgarno sequence"
FT misc_feature 6403..6591
FT /tag= i
FT /note= "Intramycelial transfer gene, ORF63"
FT misc_feature 6618..6854
FT /tag= j
FT /note= "Intramycelial transfer gene, ORF78"
FT misc_feature 6904..7608
FT /tag= k
FT /note= "Intramycelial transfer gene, ORF233"
FT misc_feature 7605..7766
FT /tag= l
FT /note= "Intramycelial transfer gene, ORF52"
FT misc_feature 7770..8009
FT /tag= m
FT /note= "Intramycelial transfer gene, ORF89"
FT RBS 8084..8088
FT /tag= n
FT /standard_name= "Shine-Dalgarno sequence"
FT misc_feature 8098..9846
FT /tag= o
FT /note= "rep gene"
FT RBS 9834..9837
FT /tag= p
FT /standard_name= "Shine-Dalgarno sequence"
FT misc_feature 9845..10055
FT /tag= q
FT /note= "Excisionase gene, xis"
FT RBS 10010..10012
FT /tag= r
FT /standard_name= "Shine-Dalgarno sequence"
FT misc_feature 10028..11188
FT /tag= s
FT /note= "Site-specific integrase gene, int"
FT
FT WO200175116-A2.
FT
FT 11-OCT-2001.
FT
FT 29-MAR-2001; 2001WO-US010321.
FT
FT 04-APR-2000; 2000US-0194461P.
FT
FT (SCHE ) SCHERING CORP.
FT
FT Hosted TJ, Horan AC;
FT
FT WPI; 2002-010791/01.
FT
FT New genes isolated from Micromonospora rosaria plasmid pMR2, useful in
FT constructing vectors for studying and expressing genes, or in
FT manipulating secondary metabolic pathways in actinomycetes.

```

```

XX PS Claim 20; Page 26-29; 34pp; English.
XX CC The invention relates to new genes isolated from Micromonospora rosaria
XX CC plasmid pMR2, and proteins encoded by such genes. The isolated genes of
XX CC of the invention are useful in the construction of vectors, which can be
XX CC used in the study and expression of genes, in manipulating secondary
XX CC metabolic pathways in actinomycetes and in creating new metabolic
XX CC products such as hybrid antibiotics. The isolated genes can also be used
XX CC to create replicating, Escherichia coli-actinomycete shuttle, integrating
XX CC and intramycelial and intramycelial conjugation vectors for use in
XX CC actinomycetes. The present sequence is M. rosaria pMR2 plasmid DNA
XX SQ Sequence 11188 BP; 1749 A; 4113 C; 3447 G; 1879 T; 0 U; 0 Other;

Query Match 4.8%; Score 56.6; DB 6; Length 11188;
Best Local Similarity 46.2%; Pred. NO. 0.18;
Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

QY 545 AGATCGGTCGGCTTATCAGCGCGCTTCCGCGCACTGGCGACCGCTCGTATGCTGCTGG 604
DB 10584 AGGTGTCGCCCTGCTGAGGCCATCAACCGCGCTACCGGACGCGGTCTCTGATCGCG 10643
QY 605 TGGCGACCGCTCTGAGTGGGTGAGCGATCGGCTCGCGCCCGCGCGGTTCGACCTGC 664
DB 10644 CTGTGTGCGGCTTACCGCGGGGAGAGATCGGAGGACTCGGGGTTCGACGTTGACCTG- 10702
QY 665 TCGCGCGCGCGCGCGCTGACCGCTGTCGAGCAGCTCCAGGAGTGGCCAGCAGCGGAG 724
DB 10703 --ACCGAACACACCATCACCGTCCGCAAGCCAGGTCGAACCGCTGCACGACAGGGGA 10760
QY 725 AGCTCGTCTTCAGTCCGCGAGACCGCGAAGCGCGCGCGCGCTGAGTTTACCACGA 784
DB 10761 AGCGTTTGAAGAGATCCCAAGTCCGAGCGCGGCAAGGAAACCATCCCATCCCTCCCC 10820
QY 785 AAGTCGCTCTACTGCTTACGCCACTCATCGCCGAAAGAAAAGTACGAGTTCGTGTTCA 844
DB 10821 ACGTGCTACCGGTGATCCGCTGACCTCGACGAGTTCGCGGGAAGGATCGCTGTTCG 10880
QY 845 CCGCGCGAAGCGGGATGTTAAGACCGCAATTCGCGCGGATTCGGTCAAGCGGT 904
DB 10881 TCAGCGCGACGATCACCCCTGCGCGGCGACACCCCTGACGAGCGTTCGACGCGTC 10940
QY 905 GCGAGGAAGCGCGCTTCGCGGCTTACGCAATTCAGCATTCGCGGACACTCACGCGGGA 964
DB 10941 GGGGAAAGTTCGACCTCGACACCCCTTCACGACCTGCGGCGACACCGGTGAGACCC 11000
QY 965 TCCTGATTCGCGCGCGCTCGCTGTCGCGCATCTCCCGCGCGCTCGTCACTGTCGA 1024
DB 11001 TCGCGCGCAGACCGGGGCGACCTTGGCGACCTGATGAGCGGCTCGGCGACTGTCGA 11060
QY 1025 TCGCGGTACGAGTCTGTATCGGCGACCTCGTCGAGAGTCCAGAGGGATTCCTCG 1084
DB 11061 TGGCTG---CGGCTCGCGGTACCTCCAGCGGTTGACGTCGTCACCGGAGATCGGCA 11117
QY 1085 CGCGGATCGAGGAGCGATGGCGCG 1111
DB 11118 AAGCCCTTCCGAGTGGCGCGCGACG 11144

RESULT 7
AAD17185/C
ID AAD17185 standard; DNA; 27541 BP.
XX
XX AAD17185;
AC
XX
XX 29-NOV-2001 (first entry)
XX Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; nys2; ds.
XX

```

```

OS Streptomyces noursei.
XX Key Location/Qualifiers
PH Complement(454..1191)
FT /*tag= a
FT /product= "NysF protein"
FT /note= "CDS does not include start codon"
FT CDS complement(1275..3092)
FT /*tag= b
FT /product= "NysG protein"
FT complement(3070..4824)
FT /*tag= c
FT /note= "CDS does not include start codon"
FT CDS /product= "NysH protein"
FT 5122..6156
FT /*tag= d
FT /product= "NysD3 protein"
FT 6338..27541
FT /*tag= e
FT /product= "NysI partial protein"
FT /note= "CDS does not include stop codon"
XX WC200159126-A2.
XX PN
XX 16-AUG-2001.
XX PF 08-FEB-2001; 2001WO-GB000509.
XX PR 08-FEB-2000; 2000GB-00002840.
XX PR 10-APR-2000; 2000GB-00008786.
XX PR 14-APR-2000; 2000GB-00009387.
XX (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE-) DZIEGLEWSKA H.
PA (ZOTC-) ZOTCHEV S B.
PA (SEKU-) SEKUROVA O N.
PA (FJAE-) FJAEVRIK E.
PA (BRAU-) BRAUTASET T.
PA (STRO-) STROM A R.
PA (VALL-) VALLA S.
XX Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX WPI; 2001-557614/62.
XX P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX as antibiotics and antifungals.
XX Claim 2; Page 151-166; 266pp; English.
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrocyclic antibiotic nystatin. The
XX nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster
XX
XX Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 U; 0 Other;
XX
XX Query Match
XX Best Local Similarity 4.7%; Score 55.8; DB 4; Length 27541;
XX Matches 346; Conservative 0; Mismatches 417; Indels 10; Gaps 3;
XX
XX 297 TGACGAGCTGACGGGAGGTCACCCAGCAGTGGGTCAACGACCTGGAGGCGGCGCGG 356
XX 4651 TGTGTGATGGGTGGTGGCTGTGTCAGATCGCGCGCGCTCGTGGGCGCGGCT 4592
XX 357 CCGTGGCGGAGTCCACGCGGGTGTGCGAAGCCGCTGCGACGAGCATACAGAA 416
XX
Db 4591 GGCCGCCCGGACCGGACGGGATGGGCGCGGATCTGCGGTCCGCGCTCTTCCGCGCAT 4532
QY 417 CTGCGACGCGCTGTGCACACGATCTGCGGCGGCGGATCGCGGCGAAACGATCAGGCT 476
Db 4531 CCGTGCATCTCCGCGCGCGGATCGCGGCTTCCGCGGCTCCCTGCTGACCGCGTC 4472
QY 477 CAACCCGCTGCTTCGACGATGCTGCCCGGCGGAGCCGAAAGAGATGAAGATTCCTGAG 536
Db 4471 GGTGAACGACGTGCAGCAGGTGCAACCTCGCCAGACCGGCTTCGCGCATCGTCTGTG 4412
QY 537 CGACCCGCGAGATCGGTGCGCTTATCACGGCGCTTCCGCGCGACTGG---CGACCGCTCGT 593
Db 4411 CGCGCGCTGATGTCCTTGGCAGGCTCTGCTGCGCGCTGCGCCAGGACGTTCCGCTGGC 4352
QY 594 CATGCTGCTGTGGCGACCGGCTCTGAGGTGGGGTGAGCCGATCGGCCCTGCGCGCGCGCG 653
Db 4351 ACTGCTGCTGTGGCGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4292
QY 654 GGTGACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 707
Db 4291 GATGGGACCGCTGTACCGCGCGGATGCAGCTGACCGCTGACCGCTGCGCGCGCGCTGCTGCG 4232
QY 708 GCTGCGCAGCAGCGGAGAGCTCGTCTTCCAGTCCCGGAAGACCGCGGCGCGCGCGCG 767
Db 4231 CGAGCGGATCACCGGGGTGGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4172
QY 768 GGTGAGTTTACACAGAAAGTCTCTACTGCTTACGCCACTCATCCCGGAAAGAAAG 827
Db 4171 CGGTTTCGCGCAGACCAACGACGCGTTCCTCGTGGTGGTGGTGGTGGTGGTGGTGGT 4112
QY 828 TGACGAGTCTGTTCACCGCGCGGAAAGCGGAGTGAAGGACGCGCAATTTCCGCGCG 887
Db 4111 CGGACGATGCTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4052
QY 888 GATCTGGGTCAAGCGCTGCGAGGAGCGCGGCTTCCGGGCTTACGCAATTCACGATCTGCG 947
Db 4051 GACCGGCTCCACCGGATCGACGCGCGCGGATGCGGATCGGCTCGCTCAGCGCGCTCT 3992
QY 948 GCACACTCAGCGCGGATCTGATTTCTGCGG---GGGCTCGCTGTGCGCGCATCTCCGCGC 1006
Db 3991 GAGTATCTCTCGCTGATCTGATGTCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3932
QY 1007 GCCTCGTCACTGCTGATCGCGCTCACGGATCTGCTGTACGGGCACTCTCGT 1059
Db 3931 GCGCGCGCGCGGCTCTGTCGCGCGGATCGCGGAGTCTCTGACACCGGCT 3879

```

RESULT 8

AAD17186/c

ID AAD17186 standard; DNA; 125401 BP.

XX AC

AAD17186;

XX 29-NOV-2001 (first entry)

XX DT

XX XX

DE Streptomyces noursei nystatin PKS gene cluster DNA.

XX XX

XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

XX XX

XX antifungal; antibiotic; ds.

XX OS

XX Streptomyces noursei.

XX XX

FH Key

CDS Location/Qualifiers

FT 6337..34771

FT /*tag= a

FT /product= "NysI complete protein"

FT 34792..51099

FT /*tag= b

FT /product= "NysJ protein"

FT 51155..57355

FT /*tag= c

FT /product= "NysK protein"

FT 57503..58687

FT CDS

```

FT      CDS      /*tag= d
FT      CDS      /product= "NysL protein"
FT      CDS      complement(58786..59980)
FT      CDS      /tag= e
FT      CDS      /product= "NysM protein"
FT      CDS      /note= "CDS does not include start codon"
FT      CDS      complement(59045..60241)
FT      CDS      /tag= f
FT      CDS      /product= "NysN protein"
FT      CDS      /note= "CDS does not include start codon"
FT      CDS      complement(60238..61296)
FT      CDS      /tag= g
FT      CDS      /product= "NysD2 complete protein"
FT      CDS      120628..121308
FT      CDS      /tag= h
FT      CDS      /product= "NysR4 (long) protein"
XX      WO200159126-A2.
XX      16-AUG-2001.
XX      08-FEB-2001; 2001WO-GB000509.
XX      08-FEB-2000; 2000GB-00002840.
XX      10-APR-2000; 2000GB-00008786.
XX      14-APR-2000; 2000GB-00009397.
XX      (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
XX      (SNTF) SINTEF STIFTTELSEN IND TEK FORSK.
XX      (ALPH-) ALPHARMA AS.
XX      (SINV-) SINVENT AS.
XX      (DZIE-) DZIEGLEWSKA H.
XX      (ZOTC-) ZOTCHEV S B.
XX      (SEKU-) SEKUROVA O N.
XX      (FJAE-) FJAEVIRK E.
XX      (BRAU-) BRAUTASET T.
XX      (STRO-) STROM A R.
XX      (VALL-) VALLA S.
XX      Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
XX      P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX      AAE10149, AAE10150.
XX      WPI: 2001-557614/62.
XX      P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
XX      AAE10149, AAE10150.
XX      New nystatin polyketide synthase polynucleotides and polypeptides, useful
XX      as antibiotics and antifungals.
XX      Claim 1; Page 188-254; 266pp; English.
XX      The present invention relates to the cloning and sequencing of the gene
XX      cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX      involved in the biosynthesis of the macrocyclic antibiotic nystatin. The
XX      nystatin PKS is useful as antifungal antibiotics. The present sequence is
XX      a Streptomyces noursei nystatin PKS gene cluster DNA
XX      SQ      Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;
XX      Query Match      4.7%; Score 55.8; DB 4; Length 125401;
XX      Best Local Similarity 44.8%; Pred. No. 0.24;
XX      Matches 346; Conservative 0; Mismatches 417; Indels 10; Gaps 3;
XX      QY      297 TGACGAGCTGGACGGGAGGTCACCCAGCAGTGGGTCAACGACCTGGAGCGCGCGCTCGG 356
XX      4651 TGTGTGATGGCGTGGTGGCGTGTGTCAGATCGCGCGCGCTCGGTGGCGCGCT 4592
XX      QY      357 CCCGTGGCGGAGTCCAGCGCGGTCTGTGGAGCCCGTGGCAGGACGATCAGCAA 416
XX      DB      4591 GGCGCGCGGACCGGACGGGATGGCGCGGATCTGGGTCCGCGCTTTCGCGCGCAT 4532
XX      QY      417 CTGCGACCGCTGCTGTCACACGATCTCGCGCGGCGGATCGCGCGGAAACGAGATCAGGCT 476

```

```

DB      4531 CCTGGACTTCTCCGCGCGGAGATCGGCGGTTTGGCACCCCTCCCTGCTGACCCGGTCTC 4472
QY      477 CAACCCGTGCTCTTCGACGATCTGCCCCGGCGGAGAGAGATGAAGTTCCTGAG 536
DB      4471 GGTGAACGACGTGACGAGGTGCAGAACCTCGCCACAGACCGGCTTCGGCATCGTCTG 4412
QY      537 CGACCCGGAGATCGGTCCGCTTATCACGGCGGTTCGCGCGCACTGG---CGACCGCTCGT 593
DB      4411 CGGCGCGCTGATGTGCTGGGACGCTGCTGTGGCGCTGCGCCAGGACGTTCCGCTGGC 4352
QY      594 CATGCTGCTGTTGGGACCGGTCTGAGGTGGGTGAGCGATCGGCTCGCGCGCGCG 653
DB      4351 ACTGCTGCTGGTGGCGCTGTGCTGTGTGGCGGTCTGCTTCGGCGCTGCTGCTGGCCCG 4292
QY      654 GGTGACCTGCTCGCGCGCGCGCGCGGCTGACCGTCTGTC-----GAGCAGTCTCAGGA 707
DB      4291 GATGGGCACTGTACGCGCGGATGACCTGACCTGGACCGGTGGGCGGCTGCTGCG 4232
QY      708 GCTGGCCAGCAGCGGAGAGCTGCTTCCACTCGCGGAGACCGGAGGCGCGCGCAC 767
DB      4231 CGAGGCGATCACCGGGGTGCGGTGCTCGGTCTTTCGTCGCGGACGACCGAGCGCGC 4172
QY      768 GGTGAGTTTCAACCAAGTTCGCTCTACTGCTTACGCCACTCATCGCGGAAAGAAAG 827
DB      4171 GCGTTTCGCGCAGACCAACGACGCGTTCCTGCTGTGTGTCGCGCGGTGGTGGCTGAT 4112
QY      828 TGACGAGTGTGTTTACCGCGCGCGGAGGCGGAGTGTAGGACGCGCAATTTCCGCGC 887
DB      4111 CGCGACGATGCTCCGCTGCTGCTGTGTGATGAACGTTTACCGTAGCCCTGCTGTG 4052
QY      888 GATCTGGTCAAGGCGTGCAGAGGAGCGCGGCTTCGCGCTTACGCATTACGATCTCGC 947
DB      4051 GACCGCTCCACCGGATCGACGCGCGCGGATCGCGCTCGCTCAGCGCGCTCT 3992
QY      948 GCACATCAGCGCGGATCTGATTTCTGCGG- GCGCTCCGCTGCGGCGATCTCCGCC 1006
DB      3991 GAGCTATCTCGCTGATCTCTGATGTCGCTGATGTCGCTGATGTCGCTTCTGCTCGGT 3932
QY      1007 GCCTCGTCACTGCTGATCGCGGTTCAGCGATCTGCTGTACGGGACCTCGCT 1059
DB      3931 GCGCGCGCGCGGTCTGTGTCGCGCGGATCGCGGAGTCTTGGACACCGGT 3879

RESULT 9
AA53491/C
ID AAX53491 standard; DNA; 114955 BP.
XX AC AAX53491;
XX DT 05-JUL-1999 (first entry)
XX DE Human adenosine A1 receptor antisense oligonucleotide fragment.
XX KW Antisense oligonucleotide; multiple target; antisense treatment;
XX KW impaired respiration; inflammation; lung disease;
XX KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
XX KW acute asthma; allergy; asthma; impeded respiration;
XX KW respiratory distress syndrome; pain; cystic fibrosis;
XX KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
XX KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
XX KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
XX KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
XX KW prostate cancer; ss.
XX OS Synthetic.
XX PN WO9913886-A1.
XX PD 25-MAR-1999.
XX PF 17-SEP-1998; 98WO-05019419.
XX PR 17-SEP-1997; 97US-0059160P.

```

PR	09-JUN-1998;	98US-00093972.
XX	(UYEC-) UNIV EAST CAROLINA.	
PA	Nyce JW;	
XX		
PI		
XX		
DR	WPI; 1999-229400/19.	
XX		
XX	New antisense oligonucleotides used in treatment of, e.g. pulmonary	
PT	vasoconstriction.	
XX		
PS	Disclosure; Page 37; 120pp; English.	
XX		
CC	The specification describes antisense oligonucleotides (AAK52869-X55271)	
CC	directed against at least 2 mRNAs selected from target genes, coding and	
CC	non-coding regions of RNAs corresponding to target genes, gene initiation	
CC	codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'	
CC	-end and the juxta-section between coding and non-coding regions and all	
CC	segments of RNAs encoding proteins associated with one or more diseases,	
CC	conditions or mixtures. The antisense oligonucleotides may be derived	
CC	from sequences AAK5272-74. These multiple target oligonucleotides	
CC	(specifically AAK5180-271) can be used for the antisense treatment of	
CC	diseases and conditions. Typical diseases and conditions are those	
CC	associated with impaired respiration and inflammation, including lung	
CC	diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,	
CC	acute asthma, allergies, asthma, impeded respiration, respiratory	
CC	distress syndrome, pain, cystic fibrosis, pulmonary hypertension,	
CC	pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary	
CC	disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.	
CC	colon cancer, breast cancer, lung cancer, pancreatic cancer,	
CC	hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as	
CC	well as all types of cancers which may metastasize or have metastasized	
CC	to the lungs, including breast and prostate cancer	
XX		
SQ	Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;	
	Query Match	4.7%; Score 55.2; DB 2; Length 114955;
	Best Local Similarity	31.6%; Pred. No. 0.32;
	Matches 298; Conservative 98; Mismatches 538; Indels 9; Gaps 4;	
QY	234 CGTGAACCTGGAGGCAACCGGATCCGCAACCCTCTGTGCCATACTCGGCATCTCAC	293
DB	105259 CGCNHNHNNNSCGGCCCGCGCGCGCGCCGCGCCNHNHNNNSCGGCCCGCGCGCGCGCC	105200
QY	294 CTTTGACGAGCTGGACGGCGAGGTTCACCCAGTAGTGGGTCAAAGCTGTGAGGCGCGGT	353
DB	105199 NNHNHNNNSCGGCCCGCGCGCGCGCCCVNHNHNNNSCGGCCCGCGCGCGCGCCV	105140
QY	354 CGGCCGCTGGCGGAGTCCAACGGGGTCTGTGGAAGCCCTGTGAGCGAAGACGATGAG	413
DB	105139 GNHNHNNNSCGGCCCGCGCGCGCGCCCGCCVGNHNHNNNSCGGCCCGCGCGCGCGCC	105080
QY	414 CAATGCCCACGGCTGTGTCAACAGATCTCGGCGCGCGCATCGCGCGGAACGATACAG	473
DB	105079 CCVGCNNHNNHNSCGGCCCGCGCGCGCGCCVGCNCNHNHNNNSCGGCCCGCGCGCC	105020
QY	474 GCTCAACCGGTGCTCTTCAGCATGTGTGCCCGCGCGAGCCGAAAGATGAATTCTT	533
DB	105019 GGCGCGCCCGCGCCCVNHNHNNNSCGGCCCGCGCGCGCCVGCNCNHNHNNNSCG	104960
QY	534 GAGCGAACCGGAGATCGTGC-GGCTTATACGGCGCTTCGCGCGACTCGCGACCGCTCG	592
DB	104959 GCCCGCGCGCGCGCGCCVCGCGVCNHNHNNNSCGGCCCGCGCGCGCGCCVCGG	104900
QY	593 TCATGCTGCTGGTGGCGACCGGTCTGAGGTGGGGTAGGCGATCGCGCTTCGCGCGCGCC	652
DB	104899 CCVGCNNHNNHNSGGCCVCGCGNNHNNHNSVGGCGVCGCGNNHNNHNSVGGCGVCGG	104840
QY	653 GGGTGCAGCTCTCTCGCGCGCGCGCCCGGTGACCGTCTGTGAGCGAGTCCAGGAGCTGG	712
DB	104839 NNNSCVGGCGVCGGNNHNNHNSCCCVGGCGVCGGNNHNNHNSGGCCVGGCGCGGNNH	104780
QY	713 CCAGCAGGGAGAGCTCTTTCAGTTCGCGGAAGACCGCGAAGCGCGCGCGCGTCA	772


```
XX Floss HG, Yu T, Leistner E;
PI XX
XX WPI; 2003-493374/46.
DR XX
XX Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Claim 7; Page 105-152; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumor agent. The present sequence is A. pretiosum
CC ansamitocin ansamitocin biosynthetic gene cluster I
XX
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;
SQ
Query Match 4.5%; Score 53; DB 7; Length 82746;
Best Local Similarity 46.6%; Pred. No. 0.8;
Matches 170; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
QY 404 AGACGATCAGCAACTGCCACCGCTGTGTCACACGATCTGCGCGGGGATCGCGCGA 463
Db 48121 AGCTGAGCCCAAGTGGCCCGGATCGCCCGCGCACTGGAGCGCGGGCGTCCGG 48062
QY 464 AACGATCAGGCTCAACCCGCTGCTTTCGACGATGTGCCCCGGCGGAGCGGAAAGAGA 523
Db 48061 ACCGATCGAGGTGCTGTGCGGACGCGCGGCGACCTCGCCGGGTGACCGGCGAC 48002
QY 524 TGAAGTCTCTGAGCGACCGGAGATCGTGGCTTATCAGCGCGCTTCGCGCACTGGC 583
Db 48001 TCGGGGCGGGTGCACCTGCTGTGTCGACGCGGACAGCGCGGTACCCCGTACT 47942
QY 584 GACCGCTGCTATGCTGCTGTCGACCGGTCTGAGTGGGTGAGGCGATCGGCTTC 643
Db 47941 ACAGCTGGCGTGCAGTGTGTCGCGCGCGGGTGGTGTGTCGTCGACACACCCCT 47882
QY 644 GCGCGCGCGGGTGCAGCTGCTCGCCGCGCGCGCGCGCTGACCGTCTGAGCAGCTCC 703
Db 47891 TCTCGGCGGGTGCAGCGACCCCGCGCGGTGACCCCGACACCGAGGCGTCCGGAC 47822
QY 704 AGGAGCTGGCAGACGCGGAGAGCTGCTTTCAGTTCGCGGAGACGCGAGGCGCGGC 763
Db 47821 TCAACCGCAGATCGCCGCGACGACCGGCTCCAGGCGGTGCTGCTCGCGCAGCGC 47762
QY 764 GCACG 768
Db 47761 GCATG 47757
RESULT 14
AAAT93095
ID AAAT93095 standard; cDNA; 24379 BP.
XX
XX AC AAAT93095;
XX
XX DT 11-MAY-1998 (first entry)
XX
XX DE Streptomyces frenolicin gene cluster.
XX
XX KW Frenolicin; antibiotic; feed additive; anticoccidial; coccidiostatic;
KW efflux pump; butyrate starter synthase; polyketide synthase; PKS;
KW hemiketalase; ketoreductase; cyclase; dehydrase; ketoreductase;
KW hydroxylase; Streptomyces roseofulvus; ds.
XX
XX OS Streptomyces sp.
XX
XX FH Key Location/Qualifiers
XX CDS 636..2948
/*tag= a
/*product= "80 kDa non-membrane protein"
FT
```

```
FT CDS
FT /note= "gene A (specifically claimed)"
FT 2945..3916
FT /*tag= b
FT /product= "membrane protein"
FT /note= "gene B (specifically claimed)"
FT 4020..4844
FT /*tag= c
FT /product= "protein with 6 membrane-spanning domains"
FT /note= "gene C (specifically claimed)"
FT 4841..6415
FT /*tag= d
FT /product= "ATP-binding component of ABC transporter"
FT /note= "gene D (specifically claimed)"
FT 6533..7183
FT /*tag= e
FT /product= "unknown non-membrane protein"
FT /note= "gene E (specifically claimed)"
FT 7344..8897
FT /*tag= f
FT /product= "putative efflux pump"
FT /note= "gene F (specifically claimed)"
FT 9164..10012
FT /*tag= g
FT /product= "transcription activator"
FT /note= "gene G (specifically claimed)"
FT complement(10105..10621)
FT /*tag= h
FT /product= "translationally coupled to gene I"
FT /note= "gene H (specifically claimed)"
FT complement(10618..11628)
FT /*tag= i
FT /product= "homologue of fabH"
FT /note= "gene I (specifically claimed)"
FT 11809..12066
FT /*tag= j
FT /product= "acyl carrier protein"
FT /note= "gene J (specifically claimed)"
FT complement(12154..13209)
FT /*tag= k
FT /product= "acyltransferase"
FT /note= "gene K (specifically claimed)"
FT 13409..14686
FT /*tag= l
FT /product= "PKS ketoacylsynthase subunit"
FT /note= "gene L (specifically claimed)"
FT 14767..16047
FT /*tag= m
FT /product= "PKS chain length factor"
FT /note= "gene M (specifically claimed)"
FT 16120..16371
FT /*tag= n
FT /product= "acyl carrier protein used by the PKS"
FT /note= "gene N (specifically claimed)"
FT complement(16453..16935)
FT /*tag= o
FT /product= "putative hemiketal dehydrase"
FT /note= "gene O (specifically claimed)"
FT 17088..17903
FT /*tag= p
FT /product= "ketoreductase related to actIII"
FT /note= "gene P (specifically claimed)"
FT 17903..18898
FT /*tag= q
FT /product= "cyclase/dehydrase related to act VII"
FT /note= "gene Q (specifically claimed)"
FT 18895..19839
FT /*tag= r
FT /product= "cyclase/dehydrase related to actIV"
FT /note= "gene R (specifically claimed)"
FT complement(19990..20907)
FT /*tag= s
FT /product= "oxidoreductase"
FT /note= "gene S (specifically claimed)"
```

ID	AAV25925 standard; cDNA, 24379 BP.	AAV25925; AAV25925	15-JUL-1998 (first entry)	Streptomyces roseofulvus frenolicin gene cluster.	Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B; antibiotic; ss.	Streptomyces roseofulvus.	Key CDS	Location/Qualifiers
XX	complement (20904. .22094)	/*tag= t					636. .2948	
XX	/*product= "quinone-forming hydroxylase"	/note= "gene T (specifically claimed)"					/*tag= a	
XX	complement (22505. .22179)	/*tag= u					/*transl_except= (pos:636. .639, aa:Met)	
XX	/*product= "not specified"	/note= "gene U"					/*note= "encodes protein given in AAW55800"	
XX	EP806480-A2.						2945. .3916	
XX	12-NOV-1997.						/*tag= b	
XX	02-MAY-1997; 97EP-00107329.						/*transl_except= (pos:2945. .2947, aa:Met)	
XX	07-MAY-1996; 96US-0016753P.						/*note= "encodes protein given in AAW55801"	
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.						4020. .4844	
XX	Reeves CD, Soliday CL;						/*tag= c	
XX	WPI; 1997-538619/50.						/*transl_except= (pos:4020. .4022, aa:Met)	
XX	P-PSDB; AAW34199, AAW34200, AAW34201, AAW34202, AAW34203, AAW34204,						/*note= "encodes protein given in AAW55802"	
XX	AAW34205, AAW34206, AAW34207, AAW34208, AAW34209, AAW34210, AAW34211,						4841. .6415	
XX	AAW34212, AAW34213, AAW34214, AAW34215, AAW34216, AAW34217, AAW34218,						/*tag= d	
XX	AAW34219.						/*note= "encodes protein given in AAW55803"	
XX	Streptomyces frenolicin gene cluster - useful for producing recombinant frenolicin antibiotics.						5533. .7183	
XX	Claim 1; Page 40-60; 66pp; English.						/*tag= e	
XX	This DNA sequence comprises the Streptomyces frenolicin gene cluster containing specifically claimed coding sequences (genes A-U) that respectively encode 21 proteins (see AAW3199-219) involved in frenolicin synthesis. The genes can be divided into 5 subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate starter synthases; (3) genes L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and cyclases/dehydrases; and (5) genes S and T encode a keto/enoyl reductase and a hydrolase. Also claimed are vectors, host cells (especially a Streptomyces sp., particularly Streptomyces roseofulvus), and the encoded proteins. Cells transformed using the above sequence can be cultured to produce frenolicins or frenolicin precursors. The precursors can be converted to frenolicins by chemical or other methods. The frenolicins can be oxidised to frenolicin B, an antibiotic used as an anticoccidial agent. The frenolicins can be used as animal feed additives						/*transl_except= (pos:6533. .6535, aa:Met)	
XX	Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;						7344. .8897	
XX	Query Match	4.5%; Score 52.8; DB 2; Length 24379;					/*tag= f	
XX	Best local similarity	52.2%; Pred. No. 0.89;					/*note= "encodes protein given in AAW55805"	
XX	Matches 117; Conservative	0; Mismatches 107; Indels 0; Gaps 0;					9164. .10012	
XX	945 GCGGCACATCAGCGCGATCTGATTTCTGCGCGCGCTCGCTCGCGATCTCCG 1004						/*tag= g	
XX	3577 GCGTACCGGAGCTGGCGGCTCGCGCGCGCGCGCGCGCGCTCTGCTCGCGG 3636						/*note= "encodes protein given in AAW55806"	
XX	1005 CGGCCTGGTCACTCGTGCATCGCGGTACGGATCTGTCAGGGCACCTCGTGAGGA 1064						/*tag= h	
XX	3637 CGCGCTCGCAACGCTCGTGTGCGCCCTGCGCTGCTCGCGCGCGCAACT 3696						/*transl_except= (pos:10621. .10619, aa:Met)	
XX	1065 GGTGACGAGGGATCTCGCGCGATCGAGGAGGATGCGCGCTCGGCTGAGGA 1124						/*note= "encodes protein given in AAW55807"	
XX	3697 CCTCGCGCGCGCTCTGCTGTGAGACCGTCTTCGCCAGGGCGGCGATCGCGCTCGC 3756						/*tag= i	
XX	1125 CCTGGAGGCGGAACCTCAGCAGGAGCTGACGGACGTGTGGCGG 1168						/*transl_except= (pos:11628. .11626, aa:Met)	
XX	3757 CATGGACGCGTCACTTACCAGGACCTGCGCGTGTGCGAGGCG 3800						/*note= "encodes protein given in AAW55808"	
XX	RESULT 15						11809. .12066	
XX	AAV25925						/*tag= j	
XX							/*note= "encodes protein given in AAW55809"	
XX							complement (12154. .13209)	
XX							/*tag= k	
XX							/*transl_except= (pos:13209. .13207, aa:Met)	
XX							/*note= "encodes protein given in AAW55810"	
XX							13409. .14686	
XX							/*tag= l	
XX							/*transl_except= (pos:13409. .13411, aa:Met)	
XX							/*note= "encodes protein given in AAW55811"	
XX							1476	

RESULT 15
AAV25925

Job time : 705.638 secs

```
FT      /tag= q
FT      /note= "encodes protein given in AAW55816"
FT      1895. .19839
FT      /tag= r
FT      /note= "encodes protein given in AAW55817"
FT      complement(1990. .20907)
FT      /tag= s
FT      /note= "encodes protein given in AAW55818"
FT      complement(20904. .22094)
FT      /tag= t
FT      /transl_except= (pos:22040. .22038,aa:Arg)
FT      /transl_except= (pos:22037. .22035,aa:Arg)
FT      /transl_except= (pos:22034. .22032,aa:Thr)
FT      /transl_except= (pos:22031. .22029,aa:lys)
FT      /note= "encodes protein given in AAW55819"
XX
XX JF10094395-A.
XX
XX
XX 14-APR-1998.
XX
XX 07-MAY-1997; 97JP-00116652.
XX
XX 07-MAY-1996; 96US-0016753P.
XX
XX 04-APR-1997; 97US-0042935P.
XX
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
XX WPI; 1998-279231/25.
XX
XX P-PSDB; AAW55800, AAW55801, AAW55802, AAW55803, AAW55804, AAW55805,
XX AAW55806, AAW55807, AAW55808, AAW55809, AAW55810, AAW55811, AAW55812,
XX AAW55813, AAW55814, AAW55815, AAW55816, AAW55817, AAW55818, AAW55819.
XX
XX Frenolicin gene cluster - useful as an antibiotic.
XX
XX
XX Claim 2; Page 36-44; 50pp: Japanese.
XX
XX The present sequence represents a frenolicin gene cluster from
XX Streptomyces roseofulvus. The present invention describes: (1) a vector
XX containing the frenolicin gene cluster DNA sequence operably connected to
XX an expression control sequence; (2) a host cell transformed by the above
XX vector; (3) a protein coded by the above DNA sequence; (4) a method for
XX the preparation of frenolicin or a biosynthetic intermediate for it in
XX which the above cell is cultured and frenolicin or its biosynthetic
XX intermediate is isolated from the culture or the cell; (5) a method for
XX the preparation of frenolicin B by oxidising frenolicin, and (6) a method
XX for the preparation of a feed composition by mixing frenolicin with other
XX components. Frenolicin B is useful as an antibiotic
XX
XX Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;
SQ
Query Match 4.5%; Score 52.8; DB 2; Length 24379;
Best Local Similarity 52.2%; Pred. No. 0.89;
Matches 117; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
Qy 945 GCGGACACTCAGCGCGGATCTGATTTCTGCGCGGCTCGCTGTCGCGGATCTCCCG 1004
Db 3577 GCGGTACCGGAGCTGGCGGCTCGCGCGCGCGCGCGCGCGCGCTGCTGCTCGCGCG 3636
Qy 1005 CCGCCTCGGTCACTGTCGATCGGTCACGATCTGTCGTACGGGACCTGCGTGAGGA 1064
Db 3637 CGCGCTCGGCAACGCTCGGTCGCGCGCTGCGCTGCTGCGGCGCACT 3696
Qy 1065 GTTCACAGGGGATCTCCGCGGATCGAGGAGGATGCGCGGCTCGGGCTGAGGA 1124
Db 3697 CCTCGCGGGGCGGTCTGTCGAGACCGTCTTCGCGAGGGCGGATCGGGCGCTCGC 3756
Qy 1125 CCTGAGCGGGAACCTCGACAGGAGCTGACGGACGTGTTGGCCG 1168
Db 3757 CATGACGCGGTACCTACCGACCTGCCCGTCGTGAGGGCG 3800
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:50:04 ; Search time 5413.76 Seconds
(without alignments)
9439.175 Million cell updates/sec

Title: US-09-855-340A-1

Perfect score: 1179

Sequence: 1 ggtgcatcgagaagaacgg.....tggcgccagcagcatga 1179

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rat.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1179	100.0	1179	6	AX338970	Sequence
2	1179	100.0	2025	7	AX150027	Bacterioph
3	1176	99.7	1176	6	AX196104	Sequence
4	66.4	5.6	347660	1	AP002994	Mesorhizo
5	65.8	5.6	300800	1	AP005036	Streptomy
6	65.6	5.6	125020	9	AF429315	Homo sapi
7	62.4	5.3	125020	9	AF429315	Homo sapi
8	61.6	5.2	2404	6	AA1401	Sequence
9	61.6	5.2	2404	7	AF030986	Mycobacte
10	60.4	5.1	2000	6	AX655393	Sequence
11	59.8	5.1	231001	2	AC130981	Rattus no
12	59.8	5.1	258319	2	AC130981	Rattus no
13	59.4	5.0	1094	11	PM7G11B	AL658196
14	59.4	5.0	3689	12	AF260831	Integrati
15	59.4	5.0	56425	7	AF165214	Bacterioph
16	57.8	4.9	2330	24	AVBHINT	Sequence
17	57.8	4.9	49220	7	AF320035	Bacterioph
18	56.6	4.8	1161	6	AR339658	Sequence
19	56.6	4.8	1161	6	AX275016	Sequence
20	56.6	4.8	11188	6	AR339644	Sequence
21	56.6	4.8	11188	6	AX275002	Sequence
22	56.6	4.8	132544	1	AF521085	Streptomy
23	56.6	4.8	231000	1	SCO939105	Streptomy
24	56.4	4.8	67480	7	AY129334	Mycobacte
25	56.4	4.8	293050	1	SCO939116	Streptomy
26	56.2	4.8	156218	2	AP004994	Oryza sat
27	55.8	4.7	27541	6	AX211706	Sequence
28	55.8	4.7	123580	1	AF263912	Streptomy
29	55.8	4.7	125401	6	AX211739	Sequence
30	55.4	4.7	65351	2	AC139773	Homo sapi
31	55.4	4.7	245476	2	AC098496	Rattus no
32	54.8	4.6	230278	14	MCU68299	Mouse cytom
33	54.2	4.6	303550	1	SCO939118	Streptomy
34	54.2	4.6	321250	1	SCO939111	Streptomy
35	53.8	4.6	1603	5	AF071875	Gallus ga
36	53.8	4.6	311000	1	SCO939122	Streptomy
37	53.6	4.5	300425	1	AP005041	Streptomy
38	53.2	4.5	2000	6	AX655393	Sequence
39	53.2	4.5	39250	1	SVI011500	Streptomy
40	53.2	4.5	132544	1	AF521085	Streptomy
41	53	4.5	575	1	MSGINT	Mycobacteri
42	53	4.5	82746	1	AF453501	Actinosyn
43	52.8	4.5	25306	1	AF058302	Streptomy
44	52.6	4.5	308050	1	SCO939124	Streptomy
45	52.2	4.4	300425	1	AP005044	Streptomy

ALIGNMENTS

RESULT 1
AX338970
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

Hosted, T.J. and Horan, A.C.
Isolation of Micromonospora carbonacea var africana plm1 integrase
and use of integrating function for site-specific integration into

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO0187936.
AX338970
AX338970.1
GI:18129106
Micromonospora carbonacea
Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

AX338970
Sequence 1 from Patent WO01

Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 1 22-NOV-2001;
SCHERING CORPORATION (US)
FEATURES
source 1. .1179
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
ORIGIN
Query Match 100.0%; Score 1179; DB 6; Length 1179;
Best Local Similarity 100.0%; Pred. No. 4.3e-159;
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGTGATCGAGAAGAACGGGCGGCTCTACCCATTCGGGACCTCTGTCGGGGTAAAAAG 60
Db 1 GTGTGATCGAGAAGAACGGGCGGCTCTACCCATTCGGGACCTCTGTCGGGGTAAAAAG 60
Qy 61 GTACCAATTCAGACCGGTTATCGAGAGAGACCGCCCAAGAAATCGGATGTGAGTTTC 120
Db 61 GTACCAATTCAGACCGGTTATCGAGAGAGACCGCCCAAGAAATCGGATGTGAGTTTC 120
Qy 121 CGTCCGAGCAGTTGACGGGCAACGGCTCATCGCGCGCGGGGCTCAGATTACCTCCGCC 180
Db 121 CGTCCGAGCAGTTGACGGGCAACGGCTCATCGCGCGCGGGGCTCAGATTACCTCCGCC 180
Qy 181 GATTTCGTGGGAGTGTGGCGGAGCTACGAAAGACGCTGAAACCGACCGCGGTGAAC 240
Db 181 GATTTCGTGGGAGTGTGGCGGAGCTACGAAAGACGCTGAAACCGACCGCGGTGAAC 240
Qy 241 TCGGAGGCAACCGGATTCGCAACCACTCTCTGCCCATCTCGGCCATCTCACCTCTTGC 300
Db 241 TCGGAGGCAACCGGATTCGCAACCACTCTCTGCCCATCTCGGCCATCTCACCTCTTGC 300
Qy 301 GAGCTGAGCGGAGTGCACCAAGAGTGGTCAACGACTGAGGCGCGGCTCGGCCGCG 360
Db 301 GAGCTGAGCGGAGTGCACCAAGAGTGGTCAACGACTGAGGCGCGGCTCGGCCGCG 360
Qy 361 TGGCGGAGTTCACGGGGGTCTCGAAGCGCTGCGGAGGAGGAGTCAAGTCAAGTTCG 420
Db 361 TGGCGGAGTTCACGGGGGTCTCGAAGCGCTGCGGAGGAGGAGTCAAGTCAAGTTCG 420
Qy 421 CAGCGCTGTGACACGAGTCTCGCGCGGGGATCGCGCGAAGCGGATCAGGCTCAAC 480
Db 421 CAGCGCTGTGACACGAGTCTCGCGCGGGGATCGCGCGAAGCGGATCAGGCTCAAC 480
Qy 481 CCGTGTCTTCGACGATCTGCCCCCGCGGAGCGGAGGAGTCAAGTCAAGTTCGAGCG 540
Db 481 CCGTGTCTTCGACGATCTGCCCCCGCGGAGCGGAGGAGTCAAGTCAAGTTCGAGCG 540
Qy 541 CCGGAGATCGTTCGGCTTATCAGGGGCTTCCCGCGCACTGCGGACCGCTCGTCATGCT 600
Db 541 CCGGAGATCGTTCGGCTTATCAGGGGCTTCCCGCGCACTGCGGACCGCTCGTCATGCT 600
Qy 601 CTGCTGCGACCGGCTCTGAGTGGGTGAGCGATCGCGCTGCGCGCGCGGCTCGAC 660
Db 601 CTGCTGCGACCGGCTCTGAGTGGGTGAGCGATCGCGCTGCGCGCGCGGCTCGAC 660
Qy 661 CTGCTGCGCGCGCGCGCGCTGACCGTCTGCGAGCAGCTCCAGGAGTGGCCAGCAG 720
Db 661 CTGCTGCGCGCGCGCGCGCGCTGACCGTCTGCGAGCAGCTCCAGGAGTGGCCAGCAG 720
Qy 721 GGAGAGTCTCTTCCAGTCCGCAAGACCGGAGGCGCGGCGACCGTCAAGTTTCAC 780
Db 721 GGAGAGTCTCTTCCAGTCCGCAAGACCGGAGGCGCGGCGACCGTCAAGTTTCAC 780
Qy 781 ACGAAGTCTCTTCTACTCTTACGCTTACGCTCATCTCGCGAAGAAAGTGCAGGTCGTG 840
Db 781 ACGAAGTCTCTTCTACTCTTACGCTTACGCTCATCTCGCGAAGAAAGTGCAGGTCGTG 840
Qy 841 TTACCGCGCGCGAAGCGCGGATGTTAAGACCGCGCAATTTCCGCGGATCTGGGTCAAG 900
Db 841 TTACCGCGCGCGAAGCGCGGATGTTAAGACCGCGCAATTTCCGCGGATCTGGGTCAAG 900

Qy 901 GCGTGCAGGAACCGGCTTCGGGCTTACGATTCAGATTCGCGCACACTACGCG 960
Db 901 GCGTGCAGGAACCGGCTTCGGGCTTACGATTCAGATTCGCGCACACTACGCG 960
Qy 961 GCGATCCTGATTTCTCGCGGGCGTCCGCTGTTCGGGATCTCCCGCGCTTCGTCCTCG 1020
Db 961 GCGATCCTGATTTCTCGCGGGCGTCCGCTGTTCGGGATCTCCCGCGCTTCGTCCTCG 1020
Qy 1021 TCGATCGCGTACGATCTGCTGACGGCACCTCGTGGAGGCTCGACGAGGGATC 1080
Db 1021 TCGATCGCGTACGATCTGCTGACGGCACCTCGTGGAGGCTCGACGAGGGATC 1080
Qy 1081 CTGCGCGGATCGAGAGGCGATGCGCGCTCGCGCTCGGGCTGAGGACCTGGAGCGGAATC 1140
Db 1081 CTGCGCGGATCGAGAGGCGATGCGCGCTCGCGCTCGGGCTGAGGACCTGGAGCGGAATC 1140
Qy 1141 GACGAGGAGTGCAGGACGCTGTTGGCGGACGCGCATGA 1179
Db 1141 GACGAGGAGTGCAGGACGCTGTTGGCGGACGCGCATGA 1179
RESULT-2
AY150027 2025 bp DNA linear PHG 02-SBP-2003
LOCUS Bacteriophage pMLP1 att/int region.
DEFINITION
ACCESSION AY150027
VERSION AY150027.1 GI:28630421
KEYWORDS Bacteriophage pMLP1
SOURCE Bacteriophage pMLP1
ORGANISM Viruses.
REFERENCE 1 (bases 1 to 2025)
AUTHORS Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and Hosted,T.J.
TITLE Development of the Micromonospora carbonacea var. africana ATCC 39149 bacteriophage pMLP1 integrase for site-specific integration in Micromonospora spp
JOURNAL Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED 12949170
REFERENCE 2 (bases 1 to 2025)
AUTHORS Hosted,T.J. Jr., Alexander,D.C. and Hewitt,D.D.
TITLE Direct Submission
JOURNAL Submitted (12-SBP-2002) New Lead Discovery, Schering-Plough Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
source 1. .2025
/organism="Bacteriophage pMLP1"
/mol_type="genomic DNA"
/specific_host="Micromonospora sp. ATCC 39149"
/db_xref="taxon:21922"
/note="host is deposited in ATCC as Micromonospora carbonacea var. africana Waitz et al."
misc_feature 1. .2025
/note="att/int region"
197..622
/note="xis"
/codon_start=1
/transl_table=11
/product="excisionase"
/protein_id="AA046044.1"
/db_xref="GI:28630422"
/translation="MRNTPGLKGTWAAVYLTARERAGLTKSELARRIKQDRTVGRW EDGKVRPDDALVARVAQLGLDLDEAALAGLRPGVTPPATPTMDLDEIELVTRDP KLDEMKRRIALLILERRERDKAAIETKRLIDLFRS"
628..1806
/gene="intX"
628..1806
/gene="intM"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="AA046045.1"
CDS
CDS
gene
CDS

```
/db_xref="GI:28630423"
/translation="MMIEKNGPVYRIRDLVRGKKVTIQGYPTKTSKXNAMVQFRAEQ
LQGNALMPRGQITTLADFYGEWMPSEYKTLKFTAVNSEGNIRNHLPIGLHLLTDEL
DGOVTQWVNDLEAGVPWPESTRGRKPLAAKTSINCHLLHTICGAAIAAKRIRLN
PCSTMLPREPEKMFLLSDPEIGRLITALPPHWRPLVLLVATGLRGEALIGLRGR
VDLLAARPLTUVVEQLQELASTGELVFQSPKAKGRRTVFTTKVALLLITPLIAGKKS
DEVFTAPKGMVRTNFRIRIWKACEAGLGLRLHDLRHHAALLISAGRPISAIS
RLGHSSIAVTDLILYGLHREEVDEGILAAIEAMAGVRAEDLEALDELTDVLADAA
"
stem_loop      join(1884..1895,1900..1911)
misc_feature    1925..1986
                 /note="attp element"
stem_loop      join(1969..1986,1991..2010)
ORIGIN
Query Match      100.0%; Score 1179; DB 7; Length 2025;
Best Local Similarity 100.0%; Pred. No. 3.8e-159;
Matches 1179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTCGTTTCGGGTAAAG 60
DB 628 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTCGTTTCGGGTAAAG 687
QY 61 GTCACATTACAGACCGGTTATCCGACGAAGACAGCGCCCAAGATCCGATGTCGAGTTC 120
DB 688 GTCACATTACAGACCGGTTATCCGACGAAGACAGCGCCCAAGATCCGATGTCGAGTTC 747
QY 121 CGTGGGACAGTTGAGGGCAACGGCGTCTATGCGCGCGGGCGGTCAAGTACCTCGCC 180
DB 748 CGTGGGACAGTTGAGGGCAACGGCGTCTATGCGCGCGGGCGGTCAAGTACCTCGCC 807
QY 181 GATTTCGTGGGAGTGGTGGCGGAGCTACGAAAGAGAGCTCAAAACCGCCCGGTGAAC 240
DB 808 GATTTCGTGGGAGTGGTGGCGGAGCTACGAAAGAGAGCTCAAAACCGCCCGGTGAAC 867
QY 241 TCGGAGGGCAACCGGATCCGAACCAACTCTCTGCCCATCTCGGCCATCTCAACCTTGC 300
DB 868 TCGGAGGGCAACCGGATCCGAACCAACTCTCTGCCCATCTCGGCCATCTCAACCTTGC 927
QY 301 GAGCTGGACGGCAGTCAACGAGAGTGGTCAACGAGCTGGAGCGCGGCGGTTCGCGCG 360
DB 928 GAGCTGGACGGCAGTCAACGAGAGTGGTCAACGAGCTGGAGCGCGGCGGTTCGCGCG 987
QY 361 TGGCGGAGTCCACGCGGCTGTGCGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 420
DB 988 TGGCGGAGTCCACGCGGCTGTGCGAAGCGCTGGCAGCGAAGACGATCAGCAACTGC 1047
QY 421 CACGCGCTCTGCACACAGTCTCGCGCGCGGCGATCGCGCGCAACCGATCAGGCTCAAC 480
DB 1048 CACGCGCTCTGCACACAGTCTCGCGCGCGGCGATCGCGCGCAACCGATCAGGCTCAAC 1107
QY 481 CCGTGCTCTTCGACGATGTCGCCCGCGCAGCGCAAGAGATGAAGTTCCTGAGCGAC 540
DB 1108 CCGTGCTCTTCGACGATGTCGCCCGCGCAGCGCAAGAGATGAAGTTCCTGAGCGAC 1167
QY 541 CCGGAGATCGTGGCTTATACCGCGGTTTCGCGCGCATGTCGCGACCGCTCGTCAATGTG 600
DB 1168 CCGGAGATCGTGGCTTATACCGCGGTTTCGCGCGCATGTCGCGACCGCTCGTCAATGTG 1227
QY 601 CTGGTGGCGACCGGCTGTAGGTGGGTGAGGGGATCGGCCTCGCGCGCGCGCGGTTCAC 660
DB 1228 CTGGTGGCGACCGGCTGTAGGTGGGTGAGGGGATCGGCCTCGCGCGCGCGCGGTTCAC 1287
QY 661 CTGCTCGCGCGCGCGCGCGCTGACCGTCTGTCGAGCAGCTCCAGGAGCTGGCCAGCAG 720
DB 1288 CTGCTCGCGCGCGCGCGCGCTGACCGTCTGTCGAGCAGCTCCAGGAGCTGGCCAGCAG 1347
QY 721 GGAGAGCTCGTCTTCCAGTTCGCGGAGACCGCGAGGGCGCGCGCAGCTTTCACC 780
DB 1348 GGAGAGCTCGTCTTCCAGTTCGCGGAGACCGCGAGGGCGCGCGCAGCTTTCACC 1407
QY 781 ACGAAGTGGCTTACTACTGCTTACCGCATCTCATCGCGGAGAAAGATGACGAGTGGT 840
```

```
DB 1408 ACGAAGTCGCTCTACTCTTACGCCACTCATCCCGGAAAGAAAGTACGAGGTGGT 1467
QY 841 TTTCACCGCGCCGAAAGCGGGATGTAAGACGCGCAATTTCCGCGCGGATCTGGGTCAAG 900
DB 1468 TTTCACCGCGCCGAAAGCGGGATGTAAGACGCGCAATTTCCGCGCGGATCTGGGTCAAG 1527
QY 901 CGGTGCGAGGAAAGCGGGCTTCGCGGCTTACGCAATTCACGATCTGCGGCACACTCACGCG 960
DB 1528 CGGTGCGAGGAAAGCGGGCTTCGCGGCTTACGCAATTCACGATCTGCGGCACACTCACGCG 1587
QY 961 CGCATCTCTGATTCTCGCGCGCGCTCCGCTGTCGCGGATCTCCCGCGCTCGGTCACTCG 1020
DB 1588 CGCATCTCTGATTCTCGCGCGCGCTCCGCTGTCGCGGATCTCCCGCGCTCGGTCACTCG 1647
QY 1021 TCGATCGCGGTTCACGAGATCTGCTAGCGGCACCTCGGTGAGGAGTTCGACGAGGGATC 1080
DB 1648 TCGATCGCGGTTCACGAGATCTGCTAGCGGCACCTCGGTGAGGAGTTCGACGAGGGATC 1707
QY 1081 CTCGCGCGGATCGAGGAGCGGATCGCGCGCTCCGCGCTGAGGACCTTGAGGCGGAACTC 1140
DB 1708 CTCGCGCGGATCGAGGAGCGGATCGCGCGCTCCGCGCTGAGGACCTTGAGGCGGAACTC 1767
QY 1141 GACGAGGAGCTGACGAGACGCTGTTGCGCGCAGCAGCATGA 1179
DB 1768 GACGAGGAGCTGACGAGACGCTGTTGCGCGCAGCAGCATGA 1806

RESULT 3
AX196104
LOCUS AX196104 1176 bp DNA linear PAT 28-AUG-2001
DEFINITION Sequence 176 from Patent WO0151639.
ACCESSION AX196104
VERSION AX196104.1 GI:15386336
KEYWORDS
SOURCE Micromonospora carbonacea
ORGANISM Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1
AUTHORS Hosted, T.J., Horan, A.C. and Wang, T.X.
TITLE Evernimicin biosynthetic genes
JOURNAL Patent: WO 0151639-A 176 19-JUL-2001;
Schering Corporation (US)
FEATURES
SOURCE 1..1176
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
1..>1176
/note="unnamed protein product"
/codon_start=1
/translation="MMIEKNGPVYRIRDLVRGKKVTIQGYPTKTSKXNAMVQFRAEQ
LQGNALMPRGQITTLADFYGEWMPSEYKTLKFTAVNSEGNIRNHLPIGLHLLTDEL
DGOVTQWVNDLEAGVPWPESTRGRKPLAAKTSINCHLLHTICGAAIAAKRIRLN
PCSTMLPREPEKMFLLSDPEIGRLITALPPHWRPLVLLVATGLRGEALIGLRGR
VDLLAARPLTUVVEQLQELASTGELVFQSPKAKGRRTVFTTKVALLLITPLIAGKKS
DEVFTAPKGMVRTNFRIRIWKACEAGLGLRLHDLRHHAALLISAGRPISAIS
RLGHSSIAVTDLILYGLHREEVDEGILAAIEAMAGVRAEDLEALDELTDVLADAA
"
CDS
Query Match      99.7%; Score 1176; DB 6; Length 1176;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 1176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTCGTTTCGGGTAAAG 60
DB 1 GTGTGGATCGAAGAACGGGCCGCTACCGCATTCGGGACCTCGTTTCGGGTAAAG 60
```


[illegible]

DLAAMLRGLVLEMAADPEGDAQASFLPAGERERQVLVWNTAYEABTASVLRREFEQ
AARTPDRAVAGHQDGLTSLYGELDVANQALHRLGLGVGAERSVLRDRGPDLLTAL
LAWKGAAPVDPVPSADRVAMEELSGAPVLTETIPVRLAISLQDDESHVAPSGPVERVA
RLAYVLTSGSTGTGKGVETVTHRGHANVMAARELASQDQCAPLFSVGAALVAVN
LWAPLVGTQAVHTVSDQMDADLGEHLVAGSPYSFKLTPGHLVDVLAQOALPEQAAAL
APVLVVAGEAFTLATLERWALAPDRLINEXGTEASVGTVHVEPRDAEDVLPJIG
RPLPNRVVYLDALQPVPGVTVGTGELVVGSTGLARGYAGRPDLTARFLPDPYGTGTPV
ARYRTGDLVRQPDGNVGFPLGRVDQVKIRGHRVELGEIQAVLTAKHPGIDAFVTV
DGLAGTYPAGAGREKREHLADRLPDYMPVATLTDALPLNANGKIDRKALPADPA
AADDDGVPARTBERTIAAQAQVLGLSERVEDGFEIGGHISRAVALVAGLRASAG
FDIGRVDFEHTVAALGHELTGRPARTVETVPARFALVPAEDARLPEDAVAYPL
TQVLQGLMELEADDTKHAVHNVSPFRITAEAPFSADALRGAAAVVAGRHEMLRTSFH
LDGFSLOPVRAHAALPVAHDLTALDEAGRRRAVEEFIERADLDLEBAPLLEW
TAHVEDGAWRLGTVCHALTEGWSHRALLMELLDYRLRGQDQSPSEAPAVRYAD
FYATEVRSLESEHRYARDVDTGCFARFALPAWAGDRSRLRQAPRETFKVAPELPPFQGL
RALAARVSLKSVLLAAHLKALSVLTEEDRFHSLGVFGARPEAFCAADRVYGVYLT
PFAPADRLTWBELVREVPFAHEAGWVEYRYPMPAIQDAGVERLLDVRFYQDPDH
VDERVVDGSGEGATEFSLAVSANDVSTYLLMTHTHALSRANADRLTALYAGVLTA
TQDPAGDAVPLPADLRLQOLVEMNDTAEATSVLRREFEQARTPRAPAVLTGE
RLTYAELDEHANRLAHLARGVGAESRVANQOLRGPVLAJLAVWKAAGAYVPVD
PSYPAERVASIVGTSAAVANVSAAYADRFAGVSLLLDDEDVSALPSAPARADLDQ
LAUVITSGTGKGVETVPHRGLAGHVAARELASRGTGGAPLFSVAFDLIIVPNL
WAPLVGTQAVFAVQDTEPAELGHRLLAGAPYSFVKLTGHLDLAEQLTTQEQELA
ELMVVAGEALPGATADRLSLDILGPRGLINEYGTFTASVGTTFVPLVAPVRDQVPIGR
PLPNMRYVLDLALQPVPGVTVGTGELVVGSTGLARGYAGRPDLTARFLPDPYAAEPGA
RLYRTGDLVRLHDPGVNVEFLGRVDDQVKIRGHRVELGEIQAVLTAKHPGIDAFVTV
PVPGRRLVGYMTPAAGACARPEELAAHCAVRLPDYMPVSAFVALEALPLNANGVKIR
GYLVPVGRGALSGTHVPRPTETKALGAINSEVLGVEEIGVHRPFDLGGHSLMI
KVLAAARAGLAVSTRMYQDHTLDLAAAVDEDTAAGARAGSQSADAEASTGAAA
ETQASGTPASAEAGSPAEAGSAGVAGSAGGARAGSAGSAGSAGSAGSAGSAGSAGS
SEBAVFTAAAAPVETPAPGVDAAPTVAVSAELLAALEQAAGGPFGRAGSGVVP
EAAALGGAAREAFVGRGDRDPLAELGTAAMAERHVPGLAVARDGAVALERGHT
LEAGSGDPVETRTLFQVGSISKHLTAFVRLVHEGLIDLDADIDGYTTSMLHADPA
APCTVTRVTLGHRAGLHRSVGRPGQLPTLLDLLEGSPVTTTPVRRELPGEV
FKKSHVTVLVQLEDVTEGPEPILMRLVLDPLGLADTSFDQHPRTAGLFLVAGH
DAHGTPLSGMRERLAAGLWNTAGDAARFVLAVRDALLGAPALVQALRAFLA
GETGTFYGLTIVDDTDLDFEGHGPESGYNMATSRMHGTFGVALTNDGSKAVV
KYLTAELGRDRTATFGSRLAADRREDPAGALPAVLAVPVVTDDEERA"
complement (11156..16285)
/note="SAV3643
nrps2 gene cluster
dihydroxy-benzoyl-glycine biosynthesis"
/codon_start=1
/transl_table=1
/product="putative non-ribosomal peptide synthetase"
/protein_id="BAC71355.1"
/db_xref="GI:29607297"
/translation="MNAELLSERFAEQRTPHAAAVVDGARTVYIELDRASLRAR
HLVDGAPETLVGLSRGVLDLIVALLAVRAGAGVPLDPAQPPARSLDLARAGA
RLVWAGAPLVDRAGVRGPEEIPDRSSDPAGLPAAADPANAAVAVFTSGTRP
AGAVVTHAGIANRIGMTVQRHALGAADRVLQTTIGFDAAGWEIIFAPLVGGTVVLA
AGARDPAALLRAVDHGVTVLVVPSVLRLLVBERGDSWGGSLRLLSAGELHAEL
VARLRETDQDLEVNVTGFTCSIDITAOQVDPALTAGVPVIGRPLPQMRVILVGN
GVPPVGVPEGLVAGGVGARGVAGRPDLTADRFPVDPVATEPGARLYRTGDQVWRS
DRTLEYGLDHQVKNVRIEPAVEEALAAHPAVTGAIVTPYADGGGKRLAALT
VSGEADPAGLRGFLAERLPDSHVSFLHTLDFPLUTANGKVDRAALPSAEIAAAGP
TYTAPRAAEELVAGVWADLLGLRDVGLDDFFALGGTSLQTLRAALRAASGEKVS
LRLGFAATVQAQGLICTPGAADGTSVAVTGGTTPAGHEGTGTCGAVHGGTGD
VDDGQTRVPTGDLPLSSGQRLWFLDRHMPGSPPEWVAPFLRLPGSLDAGTVGRA
DLAALAHPELRTVXDQGGEPQIIDAPGAGTVELRVEDTARDVAAALFGAOPERG
DLAAGPLWALLARVAGEDHVLMTMHITCDGWSTVILEREILRELCADRAGAPDL
PELQYADYGSQCARLTDVVEGELTHWRALDGLAPLDLPADHPAPVARDRGAV
VYFRVAPELAVKLGRAGHATPFMTLLTATLVARHIGQWDVPGVPGVAGTRPE
TENNVGFLNLSVLCGLTADLGTFTALRRVDRDAARALVHQEPLFELVSELPEDR
LERTPLYQVAFDLHSEGVTSVYTHSDLAAPFAWQAVAKTDLSEFMRTDPGSLDVL
KYATSLFERGTHERHTGHLTLLEQVAEQPETPLGALLSADSHRLLVWNTDAED
TDRVTFRFEQASAPARTALTGETVITYGELDAANRLAQLHRTVCVGRASVA
VLLDRPELVTTALLAVWKAAGAVVVDPSYPAERIAAMCDTAGVRTAVTTSAYAAET
SSGTRLLDLDADILAGRPAAPARTDRPRLAVTIETSGSTGTPKGVETHOGLAN
HYMAARELTPQDGGGALLSSAEDLVVPLWAPLVGTQRLFLPLQDITMSLGKRL
ADAQPFVFLKTPGHDILALQTLPAQAGALAPLVVAGEAPTRATLERWALAPDTR
LINEYGTFTASVGTTVTPIEPAHADVLPIGRPLENRYVLDPLQPVPGVAGELY
VGGTGVARGYANRPDLTADRFPLDPYGTGPGARMYRTGDLVRLHDPGNVAFVLRVDDQ

VKIRGYRVELGEIQAVLTAKHPAVRDVAVTVHVRPENGEPPTLAAHVVPADAAAPLDLAA
HCAALPPIPMIPATFTALDTTIPVNAVGRVDSALPAPORTADDESHVAPSGPVERVA
HIEWELLGQAGAHNDFFHIGNSLILRLISHLQQBFEIDFAVRTVFEQPTVARIIA
TVBERVTAQIAALSDADLLDDAARTDTTHLNNNALKEHQ"
complement (16329..16595)
/note="SAV3644
nrps2 gene cluster"
/codon_start=1
/transl_table=1
/product="putative MbH-like protein"
/protein_id="BAC71356.1"
/db_xref="GI:29607298"
/translation="MTAEQTAQQAADQPAQQAACQSGPVORVFNDEEQVSLMPAGR
DLPAWMAEAGTEGSRERELARIAEVDWDMRPLSLRRHMHSSAA"
complement (16629..17633)
/gene="arcB2"
complement (16629..17633)
/gene="arcB2"
/note="SAV3645
nrps2 gene cluster"
/codon_start=1
/transl_table=1
/product="putative ornithine cyclodeaminase"
/protein_id="BAC71357.1"
/db_xref="GI:29607299"
/translation="MLILRADVTDVLSGRETEILDLVAETRYRLHDSQTSPLHSTFL
RFPBEHRSRRIIGLPAYRGGERVAGKMWISSPFGVNAVGTDRASAAYVLSLNGH
PVAFVGEAVISAKRTAASALAARELTADNPPAAALLIGCVINLEILRFLAALPLD

	Query Match	5.6%; Score 65.8; DB 1; Length 300800;
	Best Local Similarity	48.9%; Pred. No. 0.99;
	Mismatches	280; Conservative 0; Mismatches 272; Indels 21; Gaps 3;
QY	393	GCTGGCAGCAAGACGATCACCACTGCGCAGGCTCTGTGCACACGATCTCCGGCGGCG 452
DB	11551	GGTGGGAGAGCCCGGTGTCAAGGGGTACCAACTGCTGGGGCCCTGATGAACACCGC 115492
QY	453	GATCGCGCGAAAACGGATCAGCTCAACCCGCTGCTTTCGACGATGCTGCCCGCGCGA 512
DB	115491	GGTAGATGACAGCTGATCCCGCAACCCCTGCGGATCAAAGCGGCGACATTACGA 115432
QY	513	GCCGAAGAGATGAAGTTCCTGAGCGACCCCGAGATCGTGGCTTATACGGCGCTTCC 572
DB	115431	CGTGCTGAGCGGCTGTTTTTGTGCGTGGCGGAAGTCTTTGGCGCTGCGCTATGCCATTGC 115372
QY	573	GCCGCACTGGCGACCGCTCGTCATGCTGCTGGTGGCGACCGCTCTGAGGTGGGTGAGGC 632
DB	115371	ACCGCGCTACCGGTGCTGCTCTCTCTGCTGGCTTCCAGACTCTCCGCTTCGCGGAGCT 115312
QY	633	GATCGGCTTGGCGCGCGCGGCTCGACCTGCTCGCGCGCGCGCGCGCGGCTGACCGCTGT 692
DB	115311	GGCGCGCTGCTGCGAAGGACATCGATCTGAACGGT-----CGCTTGGTTGTGGT 115261
QY	693	CGACAGCTCCAGAGCTGGCCAGCAGGAGAGCTGCTTCCAGTCGCGGAAGACCGC 752
DB	115260	CCGCGGGGCACAGCCCGAGTTGCAAGACGGCGCGCTCTTTCGACAAAGCGCCCAAGTCGCG 115201
QY	753	GAAGCGCGCGCACGCTCAGTTTTCACACGAAAGTCGCTTACTGCTTACGCCACCTCAT 812
DB	115200	CGCGGGCTTCGACCCGCTTCTTCCCGCTGAGCTGCTCGACGAGATCAGCCACCACT 115141
QY	813	-----CGCCGGAAGAAAAGTACGAGGTCTGTTTCCCGCGCGGAAAGCGCGGAT 863
DB	115140	GAAGCACTACCGCCCGCATGCTCAGGAGGGCCACTCTTCTGTTGGTCCACAGGGTGGGCA 115081
QY	864	GGAAGACGCGCAATTTCCGGCGGATCTGGGTCAAGCGCTGCGAGGAAGCGCGGCTTCC 923
DB	115080	GCITCGCGGAGCACTTCCGGGATGATCGATCAAGCCAGGACGACGCGGCGCTCTC 115021
QY	924	GGGT----TAGGATTCACGATCTGCGGCAC 953
DB	115020	GCCCAAGCTGCACTTTTCCAGCCTGAGGCATAC 114988

CDS


```

Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(36507..36887)
/genes="JPH3"
/note="synonym: JP3"
complement(36507..36887)
/genes="JPH3"
/product="junctophilin 3"
complement(36507..36887)
/genes="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17648245"
/translation="MSSGGKFNFDGGSCYGGWEDGKAHGHGVCVTPKQGGHYTGSGW
HGFEVLGVTPSGNTYQGTWAGQKHGIGLESKGVYKGEVTHGFKRGYGVRECA
NGAKYEGTWSNGLQDGYGTETYSYG"

ORIGIN
Query Match 5.3%; Score 62.4; DB 9; Length 125020;
Best Local Similarity 12.2%; Pred. No. 3.7; Mismatches 392; Indels 9; Gaps 4;
Matches 109; Conservative 387;

QY 126 GGACGAGTTCGAGGCAACCGCTCATCGCGCGCGGTGAGATTACCTCGCGCATTT 185
D 16948 GRRRSKGNWYSGMRSGSMRYTGGSXMRSSMCTSCYASMCMCWCSCCRRSCC 17007

QY 186 CGTGGGGAGTGGTGGCGAGCTACGAAAGACGCTGAACGACGCGCGTGAATCGGA 245
D 17008 MCRSYCCNRYCACCCKMSSYNTMSASYNRRSYWKRSMKWSRSGRSRCKSRGC 17067

QY 246 GGGCAACCGGATCCGCAACCACTCTCGCCATATCTCGGCCATCTCACCCCTGACGAGT 305
D 17068 GSGMKGKGGKSYGRKTKRSRKGKAKWYMYRRSMKMYSKGYCMYCWGRR 17127

QY 306 GGACGGGAGGTACCCACAGTGGGTCAACGACTGGAGCGCGCGTGGCCCGTGGCC 365
D 17128 GCYCSMTSRSAMCCSYCAKCKMCSYCYGMSKGYIYACSYRGSMSKYCMRGSTYS 17187

QY 366 GGAGTCCACGCGGGTCTGCGGAAGCGCTGGCAGGAGACGATCAGCACTGCCACGG 425
D 17188 TSCGCCCTTTTCCCCNANTGGGAGAGCTTTTNCNTKTYSYRRNGCAGC-KYNTYN 17246

QY 426 CTGTGTGACAGATCTCGGCGCGCGGATCGCGCGAAACGATCAGCTCAACCGGTG 485
D 17247 NSWRSSCRAGMCTKYCMRSMRSGNSYMYASWSRSGCYCTRCYCMWSKSKCYKYY 17306

QY 486 CTCTTCGACGATCTGCCCGCGGAGCGGAGGAGATGAAGTCTCTGAGGACCCGGA 545
D 17307 MGYMKCCWCMKCYCMRSMRSGNSYMYASWSRSGCYCTRCYCMWSKSKCYKYY 17366

QY 546 GATCGTTCGGCTTATCAGCGCGCTTCCGCGCACTGGGACCGCTCGTCACTGCTGGT 605
D 17367 MMRSKRMKGMKWSRGSWMSGASRSYCYKYSKMRSMSSKCYR--CAGNNMKGY 17424

QY 606 GCGGACCGTCTGAGTGGGGTGAAGCATCGCGCTGCGCGCGCGGGGTGACCTGCT 665
D 17425 NRYMRCSMKRRWGWKSAFYRRYWKGRGAMMCMKCYSRMSRSMCMWKKYSCASCRSCA 17484

QY 666 CGCCGCGCGCGCGCTGACCGTCTGTCGAGCAGCTCCAGGAGCTGGCCAGCAGGAGA 725

FEATURES
Db 17485 MWSGYMKCASYNWCMSSYWCCTCWCTSYMSYVRCCTCWCKGMSYTKSKWSMSSSYKGR 17544
QY 726 GCTGCTTCCAGTCCGCGAGACGCGGAGCGCGCGCGCGCGTCACTTTCACACGAA 785
D 17545 R---KSYSMCTSRGAMSCWRCCYMRGASMRAGSMRRAKGRSGWRKSWMTTGMW 17601
QY 786 AGTCGCTCTACTGCTTACGCCACTCATCGCGGAAAGAAAAGTACGAGGTCGTGTTTCC 845
D 17602 RSKYVYCTGRRMMWMTYMCWCWRRSYMYRSMAMGWRKSSWSGMMGMSASRRCKSA 17661
QY 846 CGCCCGAAAGCGGGATGTTAAGGACGCGCAATTTCCGCGCGGATCTGGGTCAAGCGTG 905
D 17662 SRSSWCSRRMKMRGSCWSSKOWGGRSRSSASCKGSRGMRSSKSSKRYKRGGRKKR 17721
QY 906 CGAGGAAGCGCGGTTCG---GGCTTACGCAATTCAGCATCTGCGGCACACTCACGCGC 962
D 17722 SMTKSGSKGSKCKWKRSGMTSSCYYSASSCMMWMSKSCMCCMMKRCACCTYSM 17781
QY 963 GATCGTGAATTTCTCGCGGCGTCTGCTGTCGGCGATCTCCGCGCGCTCGGTCACTC 1019
D 17782 SCTSMYRCMGMSYSTCGTCTCKCTGYKKSRTWYMMYSWTSKMYMWRACMMYC 17838

RESULT 8
A81401
LOCUS A81401 2404 bp DNA linear PAT 21-JAN-2000
DEFINITION Sequence 1 from Patent WO9907861.
ACCESSION A81401
VERSION A81401.1 GI:6731721
KEYWORDS
SOURCE Mycobacterium phage M66
ORGANISM Mycobacterium phage M66
REFERENCE 1 (bases 1 to 2404)
AUTHORS Da,C.G., Freitas,V.A. and Trans-Complementation,U.A.
JOURNAL Patent: WO 9907861-A 1 18-FEB-1999;
DA COSTA GARCIA MIGUEL ANGELO (PT); FREITAS VIEIRA ALCINO (PT)
FEATURES
Location/Qualifiers
1. 2404
/organism="Mycobacterium phage M66"
/mol_type="unassigned DNA"
/db_xref="taxon:65388"
697..1815
/codon_start=1
/transl_table=11
/product="INTEGRASE"
/protein_id="CAB69190.1"
/db_xref="GI:6731722"
/translation="MASVRERVRKDGTTAYLVSYRFGSGSAOACALTFDNRKAADAPA
AAYDAHAASALEWHGINPAPRGTKSELTAELVRIHDIHLTGVEQYTDKYEYLAN
DIBNLGDIPLSKLSEEDIARWYKVMETGDRDNGHAPKTLNKKYGLSGALNAAVP
RYLSTNPASGRRLPRNAEDDEIRMLTHAEFDRLDRDVAVTPHMKLMVQFVWSTLRWG
EFVFNATDGGPVRYPGFLRRVMPAVEKAGLVRPDPHDLRHTYASWQLTGGTPTTIV
SRQLGHSIQITVDVTVDVDRTSRVAEEFMDGLGDF"

Query Match 5.2%; Score 61.6; DB 6; Length 2404;
Best Local Similarity 48.5%; Pred. No. 12;
Matches 391; Conservative 0; Mismatches 364; Indels 51; Gaps 6;

QY 257 TCAGCAACCACTCTCTCCCATATCTCGCCATCTCAACCTTGACAGCTCGAGCGGCGAG 316
D 995 TCAGCAACGACATCGAACCACTCGCGACATCCCTTGTGCAAGCTCTCCGAAGAGG 1054
QY 317 TCAGCAAGAGTGGTCAACGACCTGAGCGCGCGCTCGCGCGCGTGGCGGAGTCCACGC 376
D 1055 ACATCGCGCGCTGGTGAAGTCTATGGAACCAACCGCGCGCGC-----CGACGCAACGG 1109
QY 377 GGGGTCTGCGGAAGCGCTGCGAGCGAAGACGATCAGCAACTCCACGCGCTCTGTCACA 436
```

Db 1110 GCACGCCCGGAAACCCCTCCGCAACAAATACGG---GTTCTTATCGGGGGCACTGAACGC 1166
 QY 437 CGATCTCGGGCGGGGATCGGGGAAACCGATCAGGCTCAACCGTGCTCTTCGACGA 496
 Db 1167 CGCGTCCCGCGATATTGTCACCAACCTTGGTGGCGCGCGCTGCCCGTGGGAA 1226
 QY 497 TGTGTCGCCCGCGCGAGCCGCAAGAGATGAAGTTCCTGAGACACCGGAGATCGGTCCGC 556
 Db 1227 CGCTG-----AGGACGACGACGATCGCATGCTCACCCACGCGGAGTTCGACCGGC 1279
 QY 557 TTATCAGCGGCTTCGCGCACTGCGGACCGCTCGTCTGCTGCTGCTGCGGACCGGTC 616
 Db 1280 TCCGCGACGCGGTGACACCTCACTGGAAGCTGATGCTGAGTTCAGTTCGACCGGTT 1339
 QY 617 TGAGTGGGCTGAGGCGGATCGGCTCGCGCGCGCGCGGCTCGACCTGCTCGCGCGCGGC 676
 Db 1340 TCGGTGGGTGAGGTATCGGCGCTGACGCCAGGCACTGGATTGGAGAGCTCCACGA 1399
 QY 677 CCGCGTGACCGTCTGAGACAGCTCCAGGAGCTGCGCAGCAGGAGCTCGTCTTCC 736
 Db 1400 TCAGG-----GTCCGCGACGCGCTGGAGTACTCGTCCGCGCGGTATGTTGG 1447
 QY 737 AGTCGCGGAGACGCGGAGCGGCGCGGCGACCGGTCAAGTTCACCAAGAAAGTCTCTAC 796
 Db 1448 GCGCGCGGAGACGAAACGGTCCGCGCGCACGCTGATGCGCGCCAGGTTGTTGGAGC 1507
 QY 797 TGCTTACGCCACTCATCGCGGAAAGAAAGTGAAGTGAAGTGTGTTTCAACCGCGCGAAG 856
 Db 1508 GGCTG-----GACTTGTGCAACGAGTTGTTTTCGTCAATACCGATG 1549
 QY 857 GCGGATGTGAAGACGCGCAATTCGCGCGGATCTGGTCAAGGGGTGCG-----AGG 910
 Db 1550 GTGACCGGTGAGTATCCGGGTTTCTGCTAGGTGTTGGAATCGGCTGTGGAGAGG 1609
 QY 911 AAGCGCGGCTTCGGGCTTACGCAATTCAGATTCGCGGACATCAACGCGGATCTCTGA 970
 Db 1610 CTGCTCTGGTTCGCGCGCTTACTCCGACGATTCGCGGACATCAACGCGGATCTCTGA 1669
 QY 971 TTTCTGCGGGGCTCGCTGTCGGGATCTCCGCGCGCTCGCTGCTGCTGATCGCGG 1030
 Db 1670 TAACGGCGGACACCGGTGACGATGTTGTTCTGCCAGCTGGGTGATGATGATTCAGA 1729
 QY 1031 TCAGGATCTGCTGACGGGCACTG 1056
 Db 1730 TCAGGTTGACACGTACACGATGTG 1755

RESULT 9
 AF030986 2404 bp DNA linear PHG 11-JAN-1999
 LOCUS
 DEFINITION Mycobacterium phage M56 integrase (int) gene, complete cds.
 ACCESSION AF030986
 VERSION AF030986.1 GI:2654006
 KEYWORDS
 SOURCE Mycobacterium phage M56
 ORGANISM Mycobacterium phage M56
 Viruses.
 REFERENCE 1 (bases 1 to 2404)
 Freitas-Vieira, A., Anes, E. and Moniz-Pereira, J.
 The site-specific recombination locus of mycobacteriophage M56
 determines DNA integration at the tRNA (Ala) gene of Mycobacterium
 spp
 Microbiology 144 (Pt 12), 3397-3406 (1998)
 JOURNAL
 MEDLINE 99098701
 PUBMED 9884232
 REFERENCE 2 (bases 1 to 2404)
 Vieira, A., Anes, E. and Moniz-Pereira, J.
 Mycobacterium phage M56 attP-int sequence
 TITLE Unpublished
 JOURNAL
 REFERENCE 3 (bases 1 to 2404)
 Vieira, A., Anes, E. and Moniz-Pereira, J.
 Direct Submission
 TITLE Submitted (23-OCT-1997) Microbiology, Faculty of Pharmacy
 JOURNAL

University of Lisbon, Av. Forcas Armadas, Lisbon 1600, Portugal
 Location/Qualifiers
 1. 2404
 /organism="Mycobacterium phage M56"
 /mol_type="genomic DNA"
 /db_xref="taxon:65388"
 480..502
 /rpt_type=inverted
 599..624
 /note="similar to mycobacterial common core region of
 homology between attP and attB; similar to 3' end of
 tRNA-Ala of mycobacteria; these similarities allow the
 mycobacterium tRNA-Ala gene to remain intact after
 prophage integration"
 697..1817
 /gene="int"
 697..1815
 /gene="int"
 /note="recombinase"
 /codon_start=1
 /transl_table=11
 /product="integrase"
 /protein_id="AAD03774.1"
 /db_xref="GI:2654007"
 /translation="MASVRRVRKDGTTAYLVSYRFGSGAQAALTFDNKRAADAPA
 AAVDAHGAALALEMHGINTPAPRTKSELTAIEWIRHIDHITGVEQYTDIKYQYLAN
 DIBNLGDIPLSLSEDIARVWVWETTCGRDNGHAPKTLRNKYGLSGALAAVP
 RYLSNPASGRRLPRNGAEDDRITLTHAEFDRLDAPTHWKLAVQFVMSVGLRWG
 EVSALQPHVDLETSTIRVQAKYSSAGVLPKTKSRRTVDVPALELLDLN
 EFWNTDGGPVRYPGLRRVWPAVEKAGLVERPPTPHDLRHTYASWQLTGGTPTIV
 SRQLGHESIQTIVDTYDRTSSRVAABFMDGLLGF"
 1825..1861
 /rpt_type=inverted

Query Match 5.2%; Score 61.6; DB 7; Length 2404;
 Best Local Similarity 48.5%; Pred. No. 12; Indels 51; Gaps 6;
 Matches 391; Conservative 0; Mismatches 364;
 QY 257 TCCGCAACCACTCTCGCCATCTCGGCCATCTCACCTTGACAGCTCGAGCGGCGAG 316
 Db 995 TCGCAACGACATCAACCAACCTCGCGACATCCCTTGTGGAAGCTCTCGAAGAGG 1054
 QY 317 TCACCCAGAGTGGTCAACGACCTGAGAGCGCGCGTGGCGCGGAGTCCACGC 376
 Db 1055 ACATCGCCCGCTGGGTGAAGGTCAAGAACACCGCGCGCGC-----CGACGCAACGG 1109
 QY 377 GGGGTCTGCGAAGCGCTGGCAGGAGACGATCAGCACTGCCACGCGCTGTGCACA 436
 Db 1110 GCACGCGCGGAAACCTCCCGCAACAAATAGG---GTTCCTATCGGGGCACTGAACGC 1166
 QY 437 CGATCTGCGGCGCGGAGATCGCGCGGAAAACGATCAGGCTCAACCGTGTCTTCGACGA 496
 Db 1167 CGCGCTCCCGCGATCTTGTCCACCAACCTCGCTCGGGCGCGCGCTGCGCGCGGAA 1226
 QY 497 TGCTGCGCGCGGCGGAGCGGAGAGATGAAGTTCCTGAGCAGCCCGAGATCGGTCCGC 556
 Db 1227 CGCTG-----AGGACGACGAGATCCGATCTCACCGCGGAGTTCGACCGGC 1279
 QY 557 TTATCAGCGGCTTCGCGCACTGGGACCGCTCGTCTGCTGCTGCTGCTGCGGAGCCGCTC 616
 Db 1280 TCCGCGACGCGGTGACACCTCACTGGAAGCTGATGGTTCAGTTCATGGTGTGACCGGTT 1339
 QY 617 TGAGTGGGTCAGCGGATCGGCTGCGCGCGCGCGGCTGCGCTGCTGCTGCGCGCGGC 676
 Db 1340 TCGGTGGGTGAGGTATCGGCGCTGCGAGCCAGCATGTGATTTGGAGAGCTCCACGA 1399
 QY 677 CCGCGCTGACCGTCTGTCAGCAGGTCTCCAGAGCTGGCCAGACCGGAGAGTCTGCTTCC 736
 Db 1400 TCAGG-----GTGCGGAGGCGGTGGAAGTACTCGTCCGCGCGGTATGTTGG 1447
 QY 737 AGTCGCGGAGACCGGAGGCGCGCGGCGGCGGCTGAGTTCACCAAGAAAGTCTCTAC 796

```

Db      1448  GCGCGCGGAGAGAAACGGTCCCGCCGACCGTGGATGTCGCCGCCAGGTGTTGTTGGAGC 1507
QY      797   TGTATTACGGCACTCATCGCGCGGAAAGTAAAGTACGAGTCTGTTTACCGCGCGGAAAG 856
Db      1508  GGCTG-----GACTTGTGCAACGAGTTTCTTTCGTAATACCGATG 1549
QY      857   GCGGATGTAAGGACGCGCAATTTCCGCGGATCTGGGTCAAGGGTGG-----AGG 910
Db      1550  GTGGACCGGTGAGGTATCCGGGGTTCTCGTAGGTTGGAATCCGGTGTGGAGAAAG 1609
QY      911   AAGCGGGCTTCGGGCTTACGATTACGATCTCGGCACACTCAGCGGGGATCTCTGA 970
Db      1610  CTGTTCTGTTCCGCGGCTTACTCCGACGATCTCGGCACAGTACGGTCTGGGAGC 1669
QY      971   TTTTCCGCGGCTGCGTGTGCGGGAATCTCCCGCGCTCGGTCACTCGTCAATCGCGG 1030
Db      1670  TAACGCGCGGACACCGGTGAGGATTTGTCTCGCAGCTGGTCTGAGTGGATTCAGA 1729
QY      1031  TCACGGATCTGCTGTACGGGACCTTG 1056
Db      1730  TCACGGTGACAGTACACGGATGTG 1755

RESULT 10
AX655393/c
LOCUS      2000 bp      DNA      linear      PAT 22-MAR-2003
DEFINITION Sequence 5263 from Patent WO03000898.
ACCESSION  AX655393
VERSION     AX655393.1  GI:29158207
KEYWORDS   Oryza sativa
SOURCE     Oryza sativa
ORGANISM   Oryza sativa
REFERENCE  1
AUTHORS    Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE      Plant genes involved in defense against pathogens
JOURNAL    Patent: WO 03000898-A 5263 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES   location/Qualifiers
            1..2000
            /organism="Oryza sativa"
            /mol_type="unassigned DNA"
            /db_xref="taxon:4530"

ORIGIN
Query Match      5.1%; Score 60.4; DB 6; Length 2000;
Best Local Similarity 10.7%; Pred. No.19;
Matches 65; Conservative 277; Mismatches 263; Indels 3; Gaps 1;

QY      204  GAGCTACGAAAGACGCTGAAACCGACCGCGTGAACCTCGGAGGCGAACCGATCCGCAA 263
Db      653  GARSWYKYSKSAKCKKTRTNTSYMTGMYSSYKSNWTSKWSYKWKYKWTCTWYT 594
QY      264  CCACCTCTCCCATCTCGGCATCTCATCCCTTACGAGTCTGGACGGGACGTCACCA 323
Db      593  SNKSGSTRSKMGWSGMSRMWKNWKNRKRKYMYKMKWCKTWRRCMYRWGYTMYTSR 534
QY      324  CGAGTGGGTCAAC---GACCTGGAGCGCGCTCGCGCGCGTGGCGGAGTCCACGCGGG 380
Db      533  SRMYTGRYKARYTKRRYTWYKYRYWYKYRYWYKYRYWYKYRYWYKYRYWYKYRYWY 474
QY      381  TCGTCGGAAGCGCTGGCAGCAAGACGATCATGCAACTGCCACGCGCTGCTGCACAGAT 440
Db      473  KAAYSGMWYRYKYSKNWMTSKYMSWYKCRSKYKAGCYGCKRWYVCSYGYMKW 414
QY      441  CTGCGCGCGGCGATCGCGCGGCAACGAGTACAGGTCTAGGCTCAACCGTCTCTTCGACGATGCT 500
Db      413  YTMGSYKYSRCYKYRYWYKYRYWYKYRYWYKYRYWYKYRYWYKYRYWYKYRYWYKY 354
QY      501  GCCCGCGCGGCGGCGGCAAGAGATGAAGTCTCTGAGACCGCGGAGATCGGTGCGGCTTAT 560

```

```

Db      353  GKSYKKYCTWYKMYKMRCYRWRKMRKKTKYSKRCYCWRYATCYWCCCYRKGWYSR 294
QY      561  CACGGCGCTTCGCGCGCACTGGGACCGCTCGTCACTGCTGCTGGTGGCGACCGGTCTGAG 620
Db      293  SMRTAGKWRMSWRNCRSRYWYKTKWKKSYWYKWSYKWSYKWSYKWSYKWSYKWSYK 234
QY      621  GTGGGTGAGCGCATCGGCTCGCGCGCGCGCGCGGTGCGACCTGCTCGCGCGCGCGCG 680
Db      233  STSRRAKWRACRMYSAKRRYSRTSYCGSGSYCGSSKWKYMSKSCSWMRTCSSWCSCYT 174
QY      681  GCTGACCGTCTGTCAGCAGCTCCAGAGCTGGCAGCAGCGGAGAGCTGCTCTCCAGTC 740
Db      173  CYGAGACWSCMSYMYGSCGYTRGWKRSYKYSKCKYKSCCTKYCSYTYGYYRYCKWY 114
QY      741  GCCAAGACCGCGGAGCGCGCGCGCGCGCGGTGAGTTCACCAAGAAAGTCTCTACTGCT 800
Db      113  KYSYKCYCYCYWYMYMYMCMKSCSSNMSCAYCSTSTSRWMSMYAAKWMGMC 54
QY      801  TACGCCAC 808
Db      53  GSSGMYRM 46

RESULT 11
AC130981
LOCUS      Rattus norvegicus clone CH230-229C14, WORKING DRAFT SEQUENCE.
DEFINITION AC130981
ACCESSION  AC130981
VERSION     AC130981.3  GI:30578466
KEYWORDS   HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
REFERENCE  1 (bases 1 to 231001)
AUTHORS    Muzny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
            Ayalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensheva,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,B., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwachukwu,O., Okwunodu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

```


Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinkle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 231001)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (16-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 231001)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23664662.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUVQ
Center clone name: CH230-229C14
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 216581 bases at least Q40
Consensus quality: 218902 bases at least Q30
Consensus quality: 220353 bases at least Q20
Estimated insert size: 225199; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* the accession number will be preserved.

1 231001: contig of 231001 bp in length.
Location/Qualifiers
1. .231001
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"

```

/clone="CH230-229C14"
1. .1186
/notes="wgs_contig"
complement(229055..229809)
/notes="clone boundary"
clone end:Sp6
site:EcORI
end_sequence:BZ108940"

ORIGIN
Query Match      5.1%; Score 59.8; DB 2; Length 231001;
Best Local Similarity 63.6%; Pred. No. 7.6;
Matches 91; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1029 GGTACCGATCTCTGTACGGGACCTGCTGAGGAGTTCGACGAGGGATCTCTCGGCG 1088
DB 222692 GGTGAGGAGGAGATGCGAGGAGGATGCGAGGAGGAGATGCGAGGAGGAGGAGGAGGA 222751

QY 1089 GATCGAGGAGGAGATGCGCGGCGTCCGGGCTGAGGACCTCGAGCGCGAACTCCAGCAGGA 1148
DB 222752 GATCGAGGAGGAGATGCGAGGAGGAGATGCGAGGAGGAGATGCGAGGAGGAGGAGGA 222811

QY 1149 GCTGACGAGCGTGTGCGCGAGC 1171
DB 222812 GATGAGGAGGATGATGAGGAGG 222834

RESULT 12
AC127770 258319 bp DNA linear HTG 10-OCT-2002
LOCUS Rattus norvegicus clone CH230-94P3, WORKING DRAFT SEQUENCE.
DEFINITION AC127770
ACCESSION AC127770
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION AC127770.2 GI:23665356
Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 258319)
Murny, D., Maric, M., Metzger, M., Lee, A., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., Dederich, D.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Derviz, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W.,
Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Khan, Z., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwaokemele, O., Okwuonu, G., Olarinpoogon, A., Pal, S., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

```

FEATURES
source

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodley, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.O., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

REFERENCE AUTHORS TITLE JOURNAL

COMMENT

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258319)
Direct Submission
Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 10, 2002 this sequence version replaced gi:21908142.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GXZM
Center clone name: CH230-94P3
Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 234624 bases at least Q40
Consensus quality: 237798 bases at least Q30
Consensus quality: 239372 bases at least Q20
Estimated insert size: 241391; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 258319; contig of 258319 bp in length.
* Location/Qualifiers
1..258319

FEATURES source

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-94P3"
1..1335
misc_feature
clone end:T7"
/note="wgs_end_extension
3017..68966
/note="clone_boundary
clone end:T7"
site:ECORI
end sequence:BH356710"
3050..3959
/note="clone_boundary
clone end:T7"
site:ECORI
end sequence:BH356710"
12645..13876
/note="wgs_end_extension
clone end:T7"
250432..251175
/note="clone_boundary
clone end:Sp6
site:ECORI
end sequence:BH356743"
256123..258319
/note="wgs_end_extension
clone end:T7"
ORIGIN
Query Match 5.13; Score 59.8; DB 2; Length 258319;
Best Local Similarity 63.6; Pred.No. 7.4; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 52;
QY 1029 GGTGACGAGTCTGCTGTACGGGCGACCTCGGTGAGGAGTCTGAGGGGATCTTCGGCGC 1088
DB 121339 GGTGAGGAGGAGATGAGGAGGAGATGAGGAGGAGATGAGGAGGAGATGAGGAGGAGGA 121398
QY 1089 GATGAGGAGGAGTGGCGGCTCGGCTGAGGACCTGAGGCGGAACCTGACGAGGA 1148
DB 121399 GATGAGGAGGCGGATGAGGAGGAGGATGAGGAGGAGATGAGGAGGAGATGAGGAGGAGGA 121458
QY 1149 GCTGACGAGCTGTGTTGCCGACG 1171
DB 121459 GATGAGGAGTGTGAGGAGG 121481
RESULT 13
PM7G11B/c
LOCUS PM7G11B 1094 bp DNA linear STS 29-MAY-2003
DEFINITION Penicillium marneffei STS, clone pm7g11.b, sequence tagged site.
ACCESSION AL685196
VERSION AL685196.1 GI:19337723
KEYWORDS STS.
SOURCE Penicillium marneffei
ORGANISM Penicillium marneffei
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
1
Yuen, K.Y., Pascal, G., Wong, S.S., Glaser, P., Woo, P.C., Kunst, F., Cai, J.J., Cheung, E.Y., Medigue, C. and Danchin, A.
Exploring the Penicillium marneffei genome
Arch. Microbiol. 179 (5), 339-353 (2003)
12640520
PUBMED
2 (bases 1 to 1094)
Danchin, A. and Pascal, G.
REFERENCE
AUTHORS
TITLE
Direct Submission

JOURNAL	Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre, Dextar HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
FEATURES	Location/Qualifiers
source	1. .1094 /organism="Penicillium marneffe" /mol_type="genomic DNA" /db_xref="taxon:37727" /clone="pm7g11.b"
ORIGIN	
Query Match	5.0%; Score 59.4; DB 11; Length 1094;
Best Local Similarity	43.9%; Pred. No. 31;
Matches	289; Conservative 0; Mismatches 367; Indels 2; Gaps 2;
QY	106 GCGATGGTCAGTCCGTCGGGAGAGTTCAGGAGCAACGCGCTATCGCGCGCGCGT 165
Db	838 GCGNCCCGCGCGCGTCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 779
QY	166 CAGATTACCTCGCCGATTTCGTGGGGAGTGGTGGCCGAGTACGAAAGACGTGAAA 225
Db	778 CCG 719
QY	226 CCGACCGCGTGAACTCGAGGAGCAACCGATCCGCAACACCTCTCGCCCATATCTCGC 285
Db	718 CCGCGCG-GCG 660
QY	286 CATCTCACCTTCACGAGCTGACGCGGAGTCCACCCAGCAGTGGGTCAACGACCTGGAG 345
Db	659 CCG 600
QY	346 GCGCGCGTGGCGCGTGGCGGAGTCCACGCGGGTGTTCGAGAGCCGTGCGACGCGAG 405
Db	599 GCG 540
QY	406 ACATCAGCAACTCCACGCGCTGTGTGCACACGATCTCGCGCGCGCGCGATCGCGCGAA 465
Db	539 GCG 480
QY	466 CCGATCAGGCTCAACCGTGTCTTCGAGATGCTGCCCGCGCGAGCGCAAGATG 525
Db	479 GCG 420
QY	526 AAGTTCTGACGACCGGAGATCGTCCGCTTATCACGCGCTTCGCGCGCACTGGCGA 585
Db	419 GCG 360
QY	586 CCGCTCTGCTCATGCTGTGTGGCGACCGGTCTGAGTGGGGTGAAGCGATCGGCTGCGC 645
Db	359 CCG 300
QY	646 GCGCGCGGTCGACCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 704
Db	299 CCG 240
QY	705 GCGAGTGGCGCAGCAGCGAGAGTCTCTTCAGTCCCGCGAAACCGCGCGCGCGCG 762
Db	239 GCG 182
RESULT 14	
AF260831	3689 bp DNA circular SYN 23-AUG-2000
LOCUS	
DEFINITION	Integration vector pVH-2 complete sequence.
ACCESSION	AF260831
VERSION	AF260831.1 GI:9885345
KEYWORDS	Integration vector pVH-2
SOURCE	Integration vector pVH-2
ORGANISM	artificial sequences; vectors.
REFERENCE	1 (bases 1 to 3689)
AUTHORS	Kropinski A.M., Huang V.Y.-F., Kim D. and Yung E.
TITLE	Characterization of the int-att region of bacteriophage D3, and the attD3 site on the Pseudomonas aeruginosa genome

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3689)
AUTHORS	Kropinski A.M., Huang V.Y.-F., Kim D. and Yung E.
TITLE	Direct Submission
JOURNAL	Submitted (26-APR-2000) Microbiology and Immunology, Queen's University, Room 741, Bottell Hall, Kingston, Ontario K7L 3N6, Canada
FEATURES	Location/Qualifiers
source	1. .3689 /organism="Integration vector pVH-2" /mol_type="other DNA" /db_xref="taxon:135519" /focus
source	1. .2402 /organism="Cloning vector pGEM3zf" /mol_type="other DNA" /db_xref="taxon:135522" 2402. .3689 /organism="Pseudomonas phage D3" /mol_type="other DNA" /db_xref="taxon:31535" 389. .1249 /gene="bla" 389. .1249 /gene="bla" /codon_start=1 /evidence=experimental /transl_table=11 /product="beta-lactamase" /protein_id="AAG01407.1" /db_xref="GI:9885347"
gene	/translation="MSIQHPRVALIPFFAFCIPVFAHPETLVKVKDAEDQLCARVGY IEIDLNSKILESFRPEEPFPMVSTFKVLGGVLSRIDAGQGLRRTHYSDNDVLE YSPVTEKHLTDGMTRELSAATNSNTAANLLITIGPKELTFLNMGDHYRL DRPELNEAIPNDRDITPMVAMPTTLRLKLTLGELLTLASRQQLIDWMEADKVAQPL LRGALPAGWFIADKSGAGERSGRIIAALPGDKPSRIVVIYTTGQATMDENRQIA EIGASLTKHW"
CDS	2404. .3513 /gene="int" 2404. .3513 /gene="int" /codon_start=1 /evidence=experimental /transl_table=11 /product="integrase" /protein_id="AAG01406.1" /db_xref="GI:9885348" /translation="NPKRNDISAYWISFKSATCKLVRRSSGTADYSAKALEQBR KAWKEKMGYNPPTFEVILPYLQARQHSYETTVHRIKREIFFAGRVNDLGG QDIRGYGAHRLDAGAPATINRELAALSAAINHCNTELEWALPNPKRMRREARD RWLIRAVEALCRAAGQKFGPMLEDFIRLVNTGCRREMLGLEWRRVDPANRLIYL EASHTKAGRRSIPINEGMAALKRMAFRSETSPECWVFARANGDRVLSLGAQFKQ ACQAAKATADTIHLSHTCAWLVSAGVPLADVRDLLGHSTVAMTERYAHLA PARVD AVGVLDVVRREGRSRSVHADNPALHGGPLKLVNT"
gene	2404. .3513 /gene="int"
CDS	2404. .3513 /gene="int" /codon_start=1 /evidence=experimental /transl_table=11 /product="integrase" /protein_id="AAG01406.1" /db_xref="GI:9885348" /translation="NPKRNDISAYWISFKSATCKLVRRSSGTADYSAKALEQBR KAWKEKMGYNPPTFEVILPYLQARQHSYETTVHRIKREIFFAGRVNDLGG QDIRGYGAHRLDAGAPATINRELAALSAAINHCNTELEWALPNPKRMRREARD RWLIRAVEALCRAAGQKFGPMLEDFIRLVNTGCRREMLGLEWRRVDPANRLIYL EASHTKAGRRSIPINEGMAALKRMAFRSETSPECWVFARANGDRVLSLGAQFKQ ACQAAKATADTIHLSHTCAWLVSAGVPLADVRDLLGHSTVAMTERYAHLA PARVD AVGVLDVVRREGRSRSVHADNPALHGGPLKLVNT"
ORIGIN	
Query Match	5.0%; Score 59.4; DB 12; Length 3689;
Best Local Similarity	63.8%; Pred. No. 23;
Matches	90; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY	899 AGGCGTGGGAGGACCGGGCTTCGGGCTTCCGGCTTACGATTCAGCTCGCGCACACTCACG 958
Db	3230 AGGCTGCCAGGACGAGGAAATTCGGACTTTAGATTTCAGCTTCGGGATCTCGGACACT 3289
QY	959 CGGCGATCTCTGATTTCTCGCGGGCGTCCGCTGTCCGGATCTCCCGCGCTCGGTCACT 1018
Db	3290 CGGATGGCTGGTCAGCGCGCGCTTCGGCTGGCGGATGTTCCGGATCTCTCGGACACT 3349
QY	1019 CGTGCATCGGCTCAGGATC 1039
Db	3350 CGACAGTCGGATGACTGAAC 3370

```

RESULT 15
AF165214/C
LOCUS
DEFINITION Bacteriophage D3, complete genome.
ACCESSION AF165214 L22692 U32623 U47623 AF077308 AF147978 AF165213
VERSION AF165214.1 GI:8895104
KEYWORDS
SOURCE
ORGANISM Pseudomonas phage D3
Pseudomonas phage D3
Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
Lambda-like viruses.
1 (bases 1 to 46477)
Farinha,M.A., Allan,B.J., Gertman,E.M., Ronald,S.L. and
Kropinski,A.M.
Cloning of the early promoters of Pseudomonas aeruginosa
bacteriophage D3: sequence of the immunity region of D3
J. Bacteriol. 176 (16), 4809-4815 (1994)
94327444
2 (bases 1 to 1220)
Sharp,R., Jansons,I.S., Gertman,E. and Kropinski,A.M.
Genetic and sequence analysis of the cos region of the temperate
Pseudomonas aeruginosa bacteriophage, D3
Gene 177 (1-2), 47-53 (1996)
97080499
3 (bases 53297 to 54321)
Kropinski,A.M. and Sibbald,M.J.
Transfer RNA genes and their significance to codon usage in the
Pseudomonas aeruginosa lamboid bacteriophage D3
Can. J. Microbiol. 45 (9), 791-796 (1999)
99455323
10526403
4 (bases 1 to 56425)
Gilaekjan,Z.A. and Kropinski,A.M.
Cloning and analysis of the capsid morphogenesis genes of
Pseudomonas aeruginosa bacteriophage D3: another example of protein
chain mail?
J. Bacteriol. 181 (23), 7221-7227 (1999)
20042341
10572124
5 (bases 1 to 56425)
Kropinski,A.M.
Sequence of the genome of the temperate, serotype-converting,
Pseudomonas aeruginosa bacteriophage D3
J. Bacteriol. 182 (21), 6066-6074 (2000)
20485557
11029426
6 (bases 26775 to 28246)
Kropinski,A.M., Huang,V.Y.-F., Kim,D. and Yung,E.
Characterization of the int-att region of Bacteriophage D3, and the
attB3 site on the Pseudomonas aeruginosa genome
Unpublished
7 (bases 43335 to 46477)
Farinha,M.A., Allan,B.J., Gertman,E.M., Ronald,S.L. and
Kropinski,A.M.
Direct Submission
Submitted (04-AUG-1993) Microbiology, Queen's University, Room
741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada
8 (bases 46478 to 46764)
Kropinski,A.M.
Direct Submission
Submitted (31-JUL-1995) Microbiology, Queen's University, Room
741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada
9 (bases 1 to 1220)
Kropinski,A.M. and Sharp,R.W.
Direct Submission
Submitted (29-JAN-1996) Microbiology, Queen's University, Room
741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada
10 (bases 53297 to 54321)
Kropinski,A.M. and Sibbald,M.J.
Direct Submission
Submitted (08-JUL-1998) Microbiology, Queen's University, Room
741/743, Botterell Hall, Kingston, Ontario K7L 3N6, Canada
11 (bases 1 to 7657)
Kropinski,A.M. and Gilakjan,M.A.
Direct Submission
Submitted (04-MAY-1999) Microbiology, Queen's University, Room 743,
Botterell Hall, Kingston, Ontario K7L 3N6, Canada
12 (bases 26775 to 28246)
Kropinski,A.M., Huang,V.Y.-F., Kim,D. and Yung,E.
Direct Submission
Submitted (03-JUL-1999) Microbiology, Queen's University, Room 743,
Botterell Hall, Kingston, Ontario K7L 3N6, Canada
13 (bases 1 to 56425)
Kropinski,A.M.
Direct Submission
Submitted (05-JUL-1999) Microbiology & Immunology, Queen's
University, Room 741, Botterell Hall, Kingston, Ontario K7L 3N6,
Canada
On or before Jul 2, 2000 this sequence version replaced gi:403435,
gi:984852, gi:1685011, gi:4927380, gi:5059247, gi:6166381.
FEATURES
Source
1. 56425
/organism="Pseudomonas phage D3"
/mol_type="genomic DNA"
/specific_host="Pseudomonas aeruginosa"
/db_xref="taxon:31535"
37. 417
/genes="orf1"
37. 417
/genes="orf1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Orf1"
/db_xref="GI:5059248"
/protein_id="AAD38953.1"
/translations="MARHKQPAEVAKFKGADKKNPORYOEPAKGRGDIGEAPIHLQG
PARLAKKEICTQSIKGVILGSDIILEVTANLLAEYADPTFEAVGKYTHLIGNLRL
GLTPSDQRKFGLEKPKKDEFFDF"
419. 2110
/genes="orf2"
419. 2110
/genes="orf2"
/notes="similar to Escherichia coli YmfN"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="terminase"
/protein_id="AAD38954.1"
/db_xref="GI:5059249"
/translations="MTPSDIASQYASDVVGVGAIACVYVALACOREFLNDLRQDDDDW
PYVDEAKADRAVKFQMLAPHYTKGWSASKSLVFEFQVFIENIFGVKRDITGKR
FREAYBEIPKNGKSARLAARGIYLFAADGESGAEVYSGATTEKQAFEVFRPAMWNAH
KLELNRFGEILSGNQKPMFMVMDMSKFTV:GNFGDGGASPAALVDVEHEDT
DALVDTMOTGMGAREQFLSIITTAGSNLGGPCHEKRDVIRILEGQITDITFLGIY
TIDEDDPDDPASLIKANPNYGVSVFDFDLQLOAKRSASQKAFKAGQFCYICRF
RTVMNMLAWOROKRDTTADMACRCMMALDIASKKVAALVMLFEKAGQFCYICRF
YAPAAAEENEKYQNEALEGHLVLTGSMVDYAFIADILDIAKQIDLODAFDDQA
NYILRLSNTSIPVDFPNTQVKMSDPMEVEARVIARTLWHDGPNFVMTMGVNAAK
IDAKENIYRKENDPNCKIDGFTVLIMWAGRALVAGVDDGDDFWNAIRNPPIA"
2107. 2271
/genes="orf3"
2107. 2271
/genes="orf3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Orf3"
/protein_id="AAF80770.1"
/db_xref="GI:8895112"
/translations="MNIATGLYLFVFGVLGALFVAGTFFVLGLGWLISGAASAPAIQ
AFIRKGLTSE"
2284. 3588
/genes="orf4"

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:31:14 ; Search time 141.71 Seconds
(without alignments)
4617.079 Million cell updates/sec

Title: US-09-855-340A-1

Perfect score: 1179

Sequence: 1 gtgtggtacgagagaacagg.....tgttgccgacgacgatga 1179

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56.6	4.8	1161	4	US-09-821-167-15
2	56.6	4.8	11188	4	US-09-821-167-1
3	51.4	4.4	516	4	US-09-183-861-34
4	51.4	4.4	516	4	US-09-022-765-34
5	51.4	4.4	516	4	US-09-551-974A-34
6	51.4	4.4	516	4	US-09-565-501A-34
7	51.4	4.4	516	4	US-09-639-206A-34
8	51.4	4.4	516	4	US-09-874-923-34
9	51.4	4.4	582	4	US-09-874-923-114
10	51.4	4.4	7065	4	US-09-874-923-115
11	50.4	4.3	47981	4	US-09-679-279-1
12	50.2	4.3	534	4	US-09-252-991A-11386
13	50.2	4.3	3150	4	US-09-252-991A-11487
14	50.2	4.3	3372	4	US-09-252-991A-11405
15	48.8	4.1	915	4	US-09-252-991A-2186
16	48.8	4.1	936	4	US-09-252-991A-2345
17	48.8	4.1	1059	4	US-09-252-991A-2490
18	48.6	4.1	744	4	US-09-252-991A-15538
19	48.6	4.1	807	4	US-09-252-991A-15427
20	47.4	4.0	47981	4	US-09-679-279-1
21	46.8	4.0	4403765	3	US-09-103-840A-2
22	46.8	4.0	4411529	3	US-09-103-840A-1
23	46.4	3.9	1896	4	US-09-343-011B-4
24	45.8	3.9	858	4	US-09-252-991A-12882
25	45.8	3.9	897	4	US-09-252-991A-12720
26	45.8	3.9	1398	4	US-09-252-991A-13082
27	45.8	3.9	2145	4	US-09-252-991A-13235

c 28	45.8	3.9	12001	1	US-08-458-568A-11	Sequence 11, Appl
29	45.6	3.9	1176	4	US-09-252-991A-6593	Sequence 6593, Ap
30	45.6	3.9	1440	4	US-09-252-991A-6573	Sequence 6573, Ap
31	45.6	3.9	2322	1	US-09-252-991A-6621	Sequence 6621, Ap
32	45.6	3.9	20235	1	US-07-642-734C-3	Sequence 3, Appl
33	45.6	3.9	20235	1	US-07-642-734C-3	Sequence 3, Appl
34	45.6	3.9	20235	3	US-08-439-009A-3	Sequence 3, Appl
35	45.6	3.9	20235	3	US-08-439-009A-3	Sequence 3, Appl
36	45.2	3.8	813	4	US-09-252-991A-2893	Sequence 2893, Ap
37	45.2	3.8	876	4	US-09-252-991A-2800	Sequence 2800, Ap
38	45	3.8	1296	4	US-09-252-991A-3696	Sequence 3696, Ap
39	45	3.8	1368	4	US-09-252-991A-3813	Sequence 3813, Ap
40	45	3.8	2784	4	US-09-252-991A-3769	Sequence 3769, Ap
41	45	3.8	68750	3	US-09-335-409-1	Sequence 1, Appl
42	45	3.8	68750	4	US-09-568-102-1	Sequence 1, Appl
43	45	3.8	68750	4	US-09-567-969-1	Sequence 1, Appl
44	45	3.8	68750	4	US-09-568-480-1	Sequence 1, Appl
45	45	3.8	68750	4	US-09-568-486-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-821-167-15

; Sequence 15, Application US/09821167

; Patent No. 6569668

; GENERAL INFORMATION:

; APPLICANT: Hosted Jr., Thomas J.

; APPLICANT: Horan, Ann C.

; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria

; TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom

; FILE REFERENCE: IN011490

; CURRENT APPLICATION NUMBER: US/09/821.167

; CURRENT FILING DATE: 2001-03-29

; PRIOR APPLICATION NUMBER: US 60/194,461

; PRIOR FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 15

; LENGTH: 1161

; TYPE: DNA

; ORGANISM: Micromonospora rosaria

US-09-821-167-15

Query Match 4.8%; Score 56.6; DB 4; Length 1161;

Best Local Similarity 46.2%; Pred. No. 0.0034;

Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

Qy 545 AGATCGGTGCGCTTATCAGCGCGCTTCCGCGCGACTGGCGGACCGCTGCTCATGCTGCTGG 604

Db 557 AGTCTGTCGCGCTTGTGCGGCGCATACCCGCGCTACCGGCGGCGGTCTGATCGCGG 616

Qy 605 TGGCGACCGGTCTGAGTGGGTGAGCGGATCGGCTTGGCGCGCGCGGTCGACTGC 664

Db 617 CTTGTCGCGCTTACGCGCGGAGAGATCGAGACTCCGGTTCGACAGCGTGACCTG- 675

Qy 665 TCGCGCGCGCGCGCGCTGACCGCTGTCGAGAGCTCCAGGAGTGGCGGACGCGGAG 724

Db 676 --ACCGAACACACCATCACCGTCCGCAAGCCAGGTCGAACCGCTTCACGACAGGGGGA 733

Qy 725 AGTCTGCTTTCAGTCCCGAAGACCGCGGCGCGCGCGCGCTCAGTTTCCACACGA 784

Db 734 AGCGTTTGTACAGGATCCCAAGTCCGAGGCGGCGGCAAGCAACATCGCCATCCCTCCC 793

Qy 785 AGTCTGCTTACTGCTTACGCCACTATCCCGGAAAGAAAGTACGAGGTGCTGTTC 844

Db 794 ACCTGCTACCCCGGTGATCCGCTTCACCTCGACAGTTCGCGGGAAGGATCGCTGTTCG 853

Qy 845 CCGCGCGGAAAGCGGATGTTAAGACGCGCAATTTCCGGCGGATCTGGGTCAAGGCGT 904

Db 854 TCAGCCCGAGGATCACCCCTCGCGCGGCGACACCTGTACAGGCGGTTCGTACGCGCTC 913

QY 905 GCGAGGAACCGGGCTTCGGGCTTAACGATTCAGCATTCGGGCACACTCAGCGGCGA 964
DB 914 GGGGAAGGTTCGAGACTCGACACCTCACCCTTCACGACTCGGGCACACCGGTACAGCCC 973
QY 965 TCCTGATTTCTCGCGGGCTTCGGCTGTGCGGATCTCCCGCGCTCGGTCACTGTCGA 1024
DB 974 TCGCGCGCAGACCGGGGCGACTTGGCGGACCTGATGAGCGGTTCGGGCACTGTCGA 1033
QY 1025 TCGCGGTACAGATTCGTGTACGGGACCTCGGTGAGAGGTTCAGAGGGATCTCTCG 1084
DB 1034 TGGCTG---CGGCTCGCGGTAACCTCCACGCGGTTGACGGTCTGACCGGGAGATCGCCA 1090
QY 1085 CGCGCATCGAGGAGCGATGGCGGCG 1111
DB 1091 AAGCCCTTCCGAGCTGGCGGCGCAG 1117
RESULT 2
US-09-821-167-1
; Sequence 1, Application US/09821167
; Patent No. 6549668
; GENERAL INFORMATION:
; APPLICANT: Hosted Jr., Thomas J.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria
; TITLE OF INVENTION: Plasmid pMR2 and Vectors Made Therefrom
; FILE REFERENCE: IN011490
; CURRENT APPLICATION NUMBER: US/09/821,167
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/194,461
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11188
; TYPE: DNA
; ORGANISM: Micromonospora rosaria
; FEATURE:
; NAME/KEY: RBS
; LOCATION: (6055)..(6059)
; NAME/KEY: RBS
; LOCATION: (6391)..(6394)
; NAME/KEY: RBS
; LOCATION: (8084)..(8088)
; NAME/KEY: RBS
; LOCATION: (9834)..(9837)
; NAME/KEY: RBS
; LOCATION: (10010)..(10012)
US-09-821-167-1
Query Match 4.8%; Score 56.6; DB 4; Length 11188;
Best Local Similarity 46.2%; Pred. No. 0.0043;
Matches 262; Conservative 0; Mismatches 299; Indels 6; Gaps 2;
QY 545 AGATCGGTTCGCTTATCAGCGGCTTCGCGGCACTCGCGACCGTTCGTCATGCTGCTGG 604
DB 10584 AGTCTCGCCCTGCTGAGGACCATCAACCCGGCTACCGAGCGGGTCTCTGATCGCGG 10643
QY 605 TCGCGACCGGTCTGAGGTGGGTGAGGCGATCGGCTGCGCGCGCGGGTTCGACCTGCG 664
DB 10644 CTTGTGCGGCTTACCGCGGGAGAGATCGCAGGACTCGGGTTCGACGCTGGACCTG- 10702
QY 665 TCGCGCGGCGCGCGCTGACGCTGCTGAGCAGCTCCAGGAGTGGCCACAGCGGAG 724
DB 10703 --ACCGAACACACCATCACGCTCGCGAAAGCAGGGTCAACCGCTGCACGACAGGGGA 10760
QY 725 AGCTCGTCTTCAGTCCGGAAGACCGCGAAGCGCGCGCGCTCAGTTCAGTTTCAACCA 784
DB 10761 AGCGGTTTGAAGAAGTCCCAAGTCCGAGCGGGCAAGCGAACCATCGCCATCCCTCC 10820
QY 785 AGTCTGCTTACTGCTTACGCCACTCATCGCGGAAGAAAGTACGAGGTCTGTTTCA 844
DB 10821 ACGTGTAACCGGTGATCCCGCTGACAGGTTCGCGCGGAAGGATCGCTGTTCTG 10890

QY 845 CCGCGCGAAGCGGGATGTTAAGACGCGCAATTTTCGGGGATCTGGGTCAAGGCGT 904
DB 10881 TCAGCCGCGAGGATCACCCCTCGCGGGGACACCCCTGTACAGGGTTCGTACGGCTC 10940
QY 905 GCGAGGAAGCGGGCTTCGGGCTTACGATTCACGATTCGGGCACTACACGGGCGA 964
DB 10941 GGGGAAAGTTCGAGCTCGACACCCCTCACCTTCACGACCTGGGGACACCGGTTCAGACCC 11000
QY 965 TCTGATTTCTCGCGGGCTTCGGCTGTGCGGATCTCCCGCGCTCGGTCACTGTCGA 1024
DB 11001 TCGCGCGCAGACCGGGGACCTTGGCGGACCTGATGAAGCGGCTTCGGGCACTGTCGA 11060
QY 1025 TCGCGGTACGAGTCTGCTGTACGGGCACTCGGTGAGGAGTTCGACGAGGGATCTCTCG 1084
DB 11061 TGGCTG---CGGCTCGCGGTACTCCACGCGCTTCGCGTTCGTCGCGGAGATCGCCA 11117
QY 1085 CGCGATTCGAGGAGCGATGGCGGCG 1111
DB 11118 AAGCCCTTTCGAGCTGGCGGCGCAG 11144
RESULT 3
US-09-183-861-34
; Sequence 34, Application US/09183861
; Patent No. 6365165
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,861
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/022,765
; FILING DATE: 12-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Leishmania major
US-09-183-861-34
Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;

Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
Qy 302 AGCTGGAGCGGAGGTCAACAGAGTGGGTCAACAGCTTGAGCGCGCGTTCGCGCCGT 361
Db 7 AGCAGCGCTAGACACCGCCAGCGAGCGAGCTGGAGCGCGGTGCGACGGC 66
Qy 362 GCGCGAGTCCACGCGGGGTCTCGGAGCGCGCTGGCAGCGAA-GACGATCAGCACTGC 420
Db 67 TGGCGCGGACCGCGACGAGCGCGCCAGCTGGCGCGAAACCGCGAGAGTGCAGC 126
Qy 421 CACGCGCTGTCTCACACGATCTGCGCGCGGCGATCGCGGCGAAACGATCAGGCTCAAC 480
Db 127 AGCGCTTAGACACCGCGAGCGAGCGCGCGAGCTGGAGCGCGGTGGCAGCGTGG 186
Qy 481 CCGTCTCTTTCAGCATCTGCGCGCGCGAGCGCGAGAGTGAAGTTCCTGAGCGAC 540
Db 187 CCGCGAGCGGACGAGCGCGCGCGAGCTGGCGCGAAACCGCGAGAGTGCAGCAGC 246
Qy 541 CCGGAGATCGTTCGCTTATCAGCGCGCTTCGCGCGACTGGCGACCGCTCGTCACTG 600
Db 247 GCCTAGACACCGCCAGCGAGCGCGCGAGCTGGAGCGCGAGTGCAGCGGTGGCG 306
Qy 601 CTGGTGGCGACCGGTCTGAGTGGGGTGGAGCGAT--CGGCTGCGCGCGCGCGGTGC 658
Db 307 CGAACCGCGAGAGTGCAGCGAGCGCTAGACACCGCCAGCGAGCGCGCGAGTGG 366
Qy 659 ACTGCTGCGCGCGCGCGCGCGTGCAGCGTGCAGCGAGTGCAGCGAGTGGCGAGCA 718
Db 367 AGGCGAGGTGCGAGCGGTGCGCGCGCGAGCGCGCGCGAGCGCGCGAGTGGCGCG 426
Qy 719 CGGAGAGTCTCTTCCAGTTCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCG 769
Db 427 AGCGCGAGAGTGCAGCGAGCGCTAGACACCGCGCGCGCGCGCGCGCGCGCG 477

RESULT 4
US-09-022-765-34
; Sequence 34, Application US/09022765
; Patent No. 6375955
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/022,765
; FILING DATE: 12-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.420C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Leishmania major
; US-09-022-765-34
Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
Qy 302 AGCTGGAGCGGAGGTCAACAGAGTGGGTCAACAGCTTGAGCGCGCGTTCGCGCCGT 361
Db 7 AGCAGCGCTAGACACCGCCAGCGAGCGAGCTGGAGCGCGGTGCGACGGC 66
Qy 362 GCGCGAGTCCACGCGGGGTCTCGGAGCGCGCTGGCAGCGAA-GACGATCAGCACTGC 420
Db 67 TGGCGCGGACCGCGAGCGAGCGCGCGAGCTGGCGCGAAACCGCGAGAGTGCAGC 126
Qy 421 CACGCGCTGTGACACGATCTGCGCGCGCGAGCTGGCGCGAAACGATCAGGCTCAAC 480
Db 127 AGCGCTTAGACACCGCCAGCGAGCGCGCGAGCTGGAGCGCGGTGGCAGCGTGG 186
Qy 481 CCGTCTCTTTCAGCATCTGCGCGCGCGAGCGCGAGAGTGAAGTTCCTGAGCGAC 540
Db 187 CCGCGAGCGGACGAGCGCGCGCGAGCTGGCGCGAAACCGCGAGAGTGCAGCAGC 246
Qy 541 CCGGAGATCGTTCGCTTATCAGCGCGCTTCGCGCGACTGGCGACCGCTCGTCACTG 600
Db 247 GCCTAGACACCGCCAGCGAGCGCGCGAGCTGGAGCGCGAGTGCAGCGGTGGCG 306
Qy 601 CTGGTGGCGACCGGTCTGAGTGGGGTGGAGCGAT--CGGCTGCGCGCGCGGTGC 658
Db 307 CGAACCGCGAGAGTGCAGCGAGCGCTAGACACCGCCAGCGAGCGCGCGAGTGG 366
Qy 659 ACTGCTGCGCGCGCGCGCGGTGACCGCTGCTGAGCGAGCTTCGCGAGCGCGTTCGCTG 718
Db 367 AGGCGAGGTGCGAGCGGTGCGCGCGCGAGCGCGCGCGAGCGCGCGAGTGGCGCG 426
Qy 719 CGGAGAGTCTCTTCCAGTTCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCG 769
Db 427 AGCGCGAGAGTGCAGCGAGCGCTAGACACCGCGCGCGCGCGCGCGCGCGCGCG 477

RESULT 5
US-09-551-974A-34
; Sequence 34, Application US/09551974A
; Patent No. 6500437
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C5
; CURRENT APPLICATION NUMBER: US/09/551,974A
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Leishmania major
; US-09-551-974A-34
Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
Qy 302 AGCTGGAGCGGAGGTCAACAGAGTGGGTCAACAGCTTGAGCGCGCGTTCGCGCCGT 361

7	AGCAGCGCTGTAGACACCCGCCACGACAGCGCGCCGAGCTGGAGGCACCGGTGGCACCGGC	66
362	GGCCGGAGTCCACGCGGGTGTGTGGGAAGCGCTGGCGAGCGAA-GACGATCAGCACTGC	420
67	TGGCCGCGGACCGCGACGAGCGCGCCAGCAGCTTGGCCGCGAACCGCCGAGGAGCTGCAGC	126
421	CACGCCCTGCTGCACACGATCTGGCGGGCGCGCATCGCGCGAAACGAGTACAGGCTCAAC	480
127	AGCGCTTAGACACCCCAACGACAGCGCGCCGAGCTCGAGGCGACGGGTGSCACGGCTGG	186
481	CGTGCTCTTTGACGATGCTCCCGGCGGAGCGCGAAAGAGATGAGATTCTTGAGCGAC	540
187	CCGCGGACGGCGACGAGGCCGCCACGACGCTTGGCGCGGAACGCCGAGGAGCTGCAGCAGC	246
541	CGGAGATCGTGTCGCTTATCAGCGGCTTCCGCGGCACCTGGCGCACCGCTCGTCATGCTG	600
247	GCCTAGACACCGCCACGACAGCGCGCCGAGCTGGAGGCGACAGGTGGCACCGGCTGGCGC	306
601	CTGGTGGCGACCGGCTCAGGTGGGGTGAAGCGAT--CGGCTTCGCGCGCCGCGCGGTGC	658
307	CGAAGCCGAGGAGCTGCAGCAGCGCCTAGACACCGCCACGCAAGCAGCGCGCGAGCTGG	366
659	ACCTGCTCGCCGCGCGGCCCGGTGACCGGTCTGAGCAGCTCCAGGAGCTGGCCGACGA	718
367	AGGCACGGGTGGCACGGCTGCGCGGACCGCGACGAGGCGCGCCAGCAGCTGGCCGCGA	426
719	CGGAGAGAGCTGCTTTCAGTTCGCGAAGACCGCGAAGGCCCGGCGCACGG	769
427	ACGCCGAGGAGCTGCAGCAGCGCCTTAGACACCGCCACGCGAGCAGCGCGCG	477

RESIST. 6

```

US-09-565-501A-34
; Sequence 34, Application US/09565501A
; Patent No. 6607731
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Meto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Peter Probst
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C6
; CURRENT APPLICATION NUMBER: US/09/565,501A
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Leishmania major
US-09-565-501A-34

```

Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. NO. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

302	AGCTGGA	CGGGCAGGTCTACCCAGCAGTGGTCTAA	CAGCACTTGGAGGCGGGCGCTCGGCCGT	361
7	AGCAGCGCTT	TAGACACCCGCCACGACAGCGCCGAGCTTGGAGG	CACCGGTTGGCAGCGC	66
362	GGCCGGAGTCTCACGCGGGGTCT	CGGAAGCGCTTGGCAGCGAA	-GACGATCAGCAACTGC	420
67	TGGCCGCGGACCGCACAGGCGCGCCAGCAGCTTGGCCGCGACG	CCGAGGAGCTGCAGC	126	
421	CACGGCTCTCTGCACACGATCTTGGGGCGCGCGATCGCGCGG	CGGAAA	CGGATCAGGCTCAAC	480
127	AGCGCTTAGACACGCCACGACGAGCGCGCCAGCTTGGAGG	CACCGGTTGGCAGCGCTGG	186	

QY	481	CGTGCTCTTCGACGATGCTGCCCGCGGACCGGAAAGAGATGAGTTCTCTGAGCGAC	540
DB	187	CCGCGGACGGGCGACGAGCGCCGCGACGAGTGGCCCGGAAACGCCGAGGAGCTGCGACAGC	246
QY	541	CCGGAGATCGGTCCGGCTTATCAGCGCGCTTCCGCCCGACTGGCGACCGCTGCTGATCGTG	600
DB	247	GCTTAGACACCGCCACGACGACGAGCGCGCCGAGCTGGAGGCACAGGTGGCACGCTGGCGG	306
QY	601	CTGTGTGGCGACCGGFTCTGAGTGGGGTGGAGCGAT--CGCGCTGCGCGCGCGCGCGGCTCG	658
DB	307	CGAACCGGAGGACTGCACGACGCGCTAGACACCGCCACGCAGCAGCGCGCGGAGCTGG	366
QY	659	ACCTGCTCGCCGCGCGGCCCGCGGTGACCTGCTCGAGCAGCTCCAGGAGCTGGGCCAGCA	718
DB	367	AGGACCGGGTGGCACGCGTGGCCGCGGACCGCGACGAGCGCGCCACGACAGCTGGCCGGA	426
QY	719	CGGGAGAGCTCGTCTTCCAGTGGCCGAAGACCGCGAGCGCGCGCACCG	769
DB	427	ACGCCGAGGAGCTGCAGCAGCGCGCTTAGACACCGCCACCGCAGCAGCGCGCGG	477

DECEMBER 7

```

RESOL 1 /
US-09-639-206A-34
; Sequence 34, Application US/09639206A
; Patent No. 6613337
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Meto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Skeitky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C7
; CURRENT APPLICATION NUMBER: US/09/639, 206A
; CURRENT FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Leishmania major
US-09-639-206A-34

```

Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

302	QY	AGCTGGACGGCAGGTCA	CCACAGAGTGGGTCAACACCTGGAGGCGCGCTGGCCCTT	361
7	Db	AGCAGCGCTAGACACCGCCACG	CAGCAGCGCCGAGCTGGAGGACACGGGTGGCACGCG	66
362	QY	GGCCGGAGTCCACGCGGGTCT	CGGAAGCGCGTGGCAGCGAA-GACGATCAGCAACTCC	420
67	Db	TGGCCGCGGACCGGACGAGCGCG	CCACAGATGCGCGGAGAGCTGCAGC	126
421	QY	CACGCGCTGTGCACACGATCT	GCGGCGCGCGGATCGCGCGAAACGATCAGGCTCAAC	480
127	Db	AGCGCTTAGACACCGCCACG	CAGCAGCGCGCCGAGCTGGAGGACCGGTGGCACCGCTGG	186
481	QY	CGTGCTCTTCGACGATGCTGCC	CGCGCGAGCCGAAAGATGAATTCCTGAGCCGAC	540
187	Db	CCCGGACGGCGACGAGGCCCG	CCACAGCTGCGCGCGAACGCCGAGGAGCTCCAGCAGC	246
541	QY	CCGAGAGATCGTGGCGCTTAT	CACGCGCTTCGCGCGCATCTGGCGACCGCTCGTCATGCTG	600
247	Db	GCCTAGACACCGCCACGACG	CAGCGCGCCGAGCTGGAGGACACAGGTGGCACCGCTGGCGCG	306
601	QY	CTGTGGCGGACCGGTCTTAG	GTGGGTGAGCGCAT--CGGCTTGGCGCGCGCGCGGCTGG	658

Db 307 CGAACCGGAGAGTGTGACAGCGCCCTAGACACCGCCAGCAGCGCCGAGCTGG 366
QY 659 ACTGCTCGCCGCGGCGCCGCTGACCGTCTGTGAGCAGCTCCAGAGCTGGCAGCA 718
Db 367 AGGACGGGTGACCGCTGGCGCGGACCGGACGAGCGCGCCAGCAGCTGGCGGGA 426
QY 719 CGGAGAGTCTCTTCAGTCCGACAGACCGCGAAGCGCGCGCAGG 769
Db 427 ACGCGAGGAGTGTGACGAGCGCTAGACACCGCCAGCAGCGCGCG 477

RESULT 8
US-09-874-923-34
; Sequence 34, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Leishmania major
US-09-874-923-34

Query Match 4.4%; Score 51.4; DB 4; Length 516;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGACGGGAGGTTCACCCAGCAGTGGGTCAACAGCTGGAGCGCGGCTGGCCCGT 361
Db 7 AGCAGCGCTAGACACCGCCACCGCAGCAGCGCGCGAGCTGGAGGACGGGTGGCACGGC 66
QY 362 GSCCGAGTCCAGCGGGGTCTCGGAAGCCCTGGCAGCGAA-GACGATCAGCACTGC 420
Db 67 TGGCCGCGGACCGCAGCAGCGCGCGCAGCAGTGGCGCGAAGCGCGAGGAGCTGCAGC 126
QY 421 CACGGCTGTGTGACAGTCTGCGCGCGCGGATCGCGCGGAAACGGATCAGGCTCAAC 480
Db 127 AGCGCTAGACACCGCCACCGCAGCAGCGCGCGAGCTGGAGGACGGGTGGCAGCTGG 186
QY 481 CGTGTCTTTCAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 187 CGCGGACCGCGCAGGCG 246
QY 541 CGGAGATCGGTGGCTTATCAGCGCGCTTCCGCGCGCTGCGCGCGCGCGCGCGCGCG 600
Db 247 GCTTAGACCGCCACCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
QY 601 CTGTTGGCAGCGGTCTGAGGTGGGTGAGGCGAT--CGGCTTGGCGCGCGCGCGCGCG 658
Db 307 CGAACCGGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCGCGCTGG 366
QY 659 ACTGCTCGCGCGCGCGCGCGCGCTGACCGTCTGTGAGCAGCTCCAGGAGCTGGCAGCA 718
Db 367 AGGACGGGTGACCGCTGGCGCGGACCGCGACGAGCGCGCGCGCGCGCGCGCGCGCG 426
QY 719 CGGAGAGTCTGCTTCTCCAGTCCCGAAGACCGCGAAGCGCGCGCGCGCGCGCG 769

Db 427 ACGCGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCG 477

RESULT 9
US-09-874-923-114
; Sequence 114, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Leishmania major and chagasi
US-09-874-923-114

Query Match 4.4%; Score 51.4; DB 4; Length 582;
Best Local Similarity 47.1%; Pred. No. 0.036;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGACGGGAGGTTCACCCAGCAGTGGGTCAACAGCTGGAGCGCGGCTGGCCCGT 361
Db 59 AGCAGCGCTAGACACCGCCACCGCAGCAGCGCGCGAGCTGGCGCGAAGCGCGAGGTGCAGC 118
QY 362 GSCCGAGTCCAGCGGGGTCTCGGAAGCCCTGGCAGCGAA-GACGATCAGCACTGC 420
Db 119 TGGCCGCGGACCGCAGCAGCGCGCGCAGCAGTGGCGCGAAGCGCGAGGAGTGCAGC 178
QY 421 CACGGCTGTGTGACAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Db 179 AGCGCTTAGACACCGCCACCGCAGCAGCGCGCGCGAGCTGGAGGACGGGTGGCAGCGTGG 238
QY 481 CGGTGTCTTTCAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 239 CGCGGAGCGCGAGGCG 298
QY 541 CGGAGATCGGTGGCTTATCAGCGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db 299 GCCTAGACACCGCCACCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 358
QY 601 CTGTTGGCAGCGGTCTGAGTGGGTGAGGCGAT--CGGCTTGGCGCGCGCGCGCGCGTGG 658
Db 359 CGAACCGGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCGCGCTGG 418
QY 659 ACTGCTCGCGCGCGCGCGCGCGTGTGAGCAGCTCCAGGAGCTGGCAGCA 718
Db 419 AGGACGGGTGGCAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
QY 719 CGGAGAGTCTGCTTCTCCAGTCCCGAAGACCGCGAAGCGCGCGCGCGCGCGCG 769
Db 479 ACGCGAGGAGTGTGACAGCGCTAGACACCGCCAGCAGCGCGCGCGCGCGCG 529

RESULT 10
US-09-874-923-115
; Sequence 115, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Bhatia, Ajay
APPLICANT: Coler, Rhea
APPLICANT: Probst, Peter
APPLICANT: Brannon, Mark
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
FILE REFERENCE: 210121.420C8
CURRENT APPLICATION NUMBER: US/09/874,923
CURRENT FILING DATE: 2001-06-04
NUMBER OF SEQ ID NOS: 122
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 115
LENGTH: 7065
TYPE: DNA
ORGANISM: Leishmania major and chagasi
US-09-874-923-115

Query Match 4.4%; Score 51.4; DB 4; Length 7065;
Best Local Similarity 47.1%; Pred. No. 0.048;
Matches 222; Conservative 0; Mismatches 246; Indels 3; Gaps 2;

QY 302 AGCTGACGGGAGGTACCCAGCAGTGGTCAACGACCTGGAGCGCGCTCGCCCGT 361
Db |||||
QY 1283 AGCAGCGCTTAGACACCGCCAGCAGCGCGCGCTGGAGGACCGGTGGCAGCGC 1342
Db |||||
QY 362 GCGCGAGTCCAGCGCGGTGTCGGAAGCGGTGTCAGCGAA-GACGATCAGCAATGC 420
Db |||||
QY 1343 TGGCCGGACCGCAGGAGCGCGCAGCAGCTGCGCGGAACCGAGGAGTGCAGC 1402
Db |||||
QY 421 CAGGCGCTGTGACACAGATCTCGCGCGCGGATCGCGCGAAGCGATCAGGCTCAAC 480
Db |||||
QY 1403 AGCGCTTAGACACCGCCAGCAGCAGCGCGCGAGCTGGAGGACCGGTGGCAGCGCTGG 1462
Db |||||
QY 481 CCGTGTCTTCGACGATGTGCCCCGCGAGCGCGAAGAGATGAAGTTCCTGAGCGAC 540
Db |||||
QY 1463 CCGCGACCGGACGAGGCGCGCCAGCAGCTGCGCGCGAAGCGAGCTGCGAGCAGC 1522
Db |||||
QY 541 CCGAGATCGTGGCTTATCAGCGCGCTTCGCGCGCATGCGCGACCGCTGTCATGCTG 600
Db |||||
QY 1523 GCCTAGACACCGCCAGCAGCGCGCGAGCTGGAGGACACAGTGGCAGCGCTGGCGC 1582
Db |||||
QY 601 CTGCTGCGACCGTCTGAGGTGGGTGAGCGAT--CGGCTGCGCGCGCGCGGTGCG 658
Db |||||
QY 1583 CGAACCGGAGGAGTGTGAGCAGCGCTAGACACCGCCACGAGCAGCGCGCGAGCTGG 1642
Db |||||
QY 659 ACCTGTGCGCGCGCGCGCGCTGACCGTGTGTCAGCAGCTCCAGGAGCTGGCGCAGCA 718
Db |||||
QY 1643 AGCAGCGGTGGCAGCGCTGGCGCGGACCGGAGCGCGCGCAGCAGCTGGCGCGCA 1702
Db |||||
QY 719 CGGAGAGCTGCTTCCAGTGCCTCCAGACCGCGAGCGCGCGCGCGCG 769
Db |||||
QY 1703 ACGCCGAGGAGTGTGACGAGCGCTAGACACCGCCACGAGCAGCGCGCGCG 1753
Db |||||

RESULT 11
US-09-679-279-1
Sequence 1, Application US/09679279
Patent No. 6524841
GENERAL INFORMATION:
APPLICANT: McDaniel, Robert
APPLICANT: Volchegursky, Yanina
TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
TITLE OF INVENTION: Genes and Uses Thereof
FILE REFERENCE: 30622004700
CURRENT APPLICATION NUMBER: US/09/679,279
CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158,305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190,024

PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (928)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCIII homolog),
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acyltransferase
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDII, deoxysugar transaminase (eryCI, DnrJ homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (5822)...(6595)
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU hom
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megEII-1 (megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI; SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc_feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc_feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc_feature
LOCATION: (13849)...(15126)
OTHER INFORMATION: megAI, KS1
NAME/KEY: misc_feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megAI, AT1

```
; NAME/KEY: misc_feature
; LOCATION: (17155)...(17694)
; OTHER INFORMATION: megAI, KR1
; NAME/KEY: misc_feature
; LOCATION: (17947)...(18207)
; OTHER INFORMATION: megAI, ACP1
; NAME/KEY: misc_feature
; LOCATION: (18288)...(19548)
; OTHER INFORMATION: megAI, KS2
; NAME/KEY: misc_feature
; LOCATION: (19876)...(20910)
; OTHER INFORMATION: megAI, AT2
; NAME/KEY: misc_feature
; LOCATION: (21517)...(22053)
; OTHER INFORMATION: megAI, KR2
; NAME/KEY: misc_feature
; LOCATION: (22318)...(22575)
; OTHER INFORMATION: megAI, ACP2
; NAME/KEY: CDS
; LOCATION: (22867)...(33555)
; OTHER INFORMATION: megAI; SEQ ID NO: 14= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (22957)...(24237)
; OTHER INFORMATION: megAI, KS3
; NAME/KEY: misc_feature
; LOCATION: (24544)...(25581)
; OTHER INFORMATION: megAI, AT3
; NAME/KEY: misc_feature
; LOCATION: (26230)...(26733)
; OTHER INFORMATION: megAI, KR3 (inactive)
; NAME/KEY: misc_feature
; LOCATION: (26938)...(27258)
; OTHER INFORMATION: megAI, ACP3
; NAME/KEY: misc_feature
; LOCATION: (27393)...(28590)
; OTHER INFORMATION: megAI, KS4
; NAME/KEY: misc_feature
; LOCATION: (28897)...(29931)
; OTHER INFORMATION: megAI, AT4
; NAME/KEY: misc_feature
; LOCATION: (29953)...(30477)
; OTHER INFORMATION: megAI, DH4
; NAME/KEY: misc_feature
; LOCATION: (31396)...(32244)
; OTHER INFORMATION: megAI, ER4
; NAME/KEY: misc_feature
; LOCATION: (32257)...(32799)
; OTHER INFORMATION: megAI, KR4
; NAME/KEY: misc_feature
; LOCATION: (33052)...(33312)
; OTHER INFORMATION: megAI, ACP4
; NAME/KEY: CDS
; LOCATION: (33666)...(43271)
; OTHER INFORMATION: megAI; SEQ ID NO: 15= translated amino acid sequence
; NAME/KEY: misc_feature
; LOCATION: (33780)...(35027)
; OTHER INFORMATION: megAI, KS5
; NAME/KEY: misc_feature
; LOCATION: (35385)...(36419)
; OTHER INFORMATION: megAI, AT5
; NAME/KEY: misc_feature
; LOCATION: (37068)...(37604)
; OTHER INFORMATION: megAI, KR5
; NAME/KEY: misc_feature
; LOCATION: (37860)...(38120)
; OTHER INFORMATION: megAI, ACP5
; NAME/KEY: misc_feature
; LOCATION: (38187)...(39470)
; OTHER INFORMATION: megAI, KS6
; NAME/KEY: misc_feature
; LOCATION: (39795)...(40811)
; OTHER INFORMATION: megAI, AT6
; NAME/KEY: misc_feature
```

```
; LOCATION: (41406)...(41936)
; OTHER INFORMATION: megAI, KR6
; NAME/KEY: misc_feature
; LOCATION: (42168)...(42425)
; OTHER INFORMATION: megAI, ACP6
; NAME/KEY: misc_feature
; LOCATION: (42585)...(43271)
; OTHER INFORMATION: megAI, TE
; NAME/KEY: CDS
; LOCATION: (43268)...(44344)
; OTHER INFORMATION: megAI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
; OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (44355)...(45623)
; OTHER INFORMATION: megAI, desosaminyl transferase, desosamine glycosyltransferase;
; OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (45620)...(46591)
; OTHER INFORMATION: megAI-2 (megBI), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
; OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
; OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (46660)...(47403)
; OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence
; NAME/KEY: CDS
; LOCATION: (47411)...(47980)
; OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence
; US-09-679-279-1
;
; Query Match 4.3%; Score 50.4; DB 4; Length 47981;
; Best Local Similarity 46.3%; Pred. No. 0.096;
; Matches 251; Conservative 0; Mismatches 276; Indels 15; Gaps 2;
;
Qy 219 GCTGAACCGACCGCGTGAACTCGGAGGACCAACCGGATCCGCAACCACTCTCTGCCCAT 278
Db 29721 GCTGATGGCGTTCGAGGAGACCGCCGAGAGCGCGCGCTCACCGCGTCCGAC 29780
Qy 279 ACTCGGCATCTACCTTTCAGAGCTGAGCGGAGGTACCCAGAGTGGTCAACGA 338
Db 29781 GCTCGCCGCGACCAACGACGGCGCTGCGGATTTCTCCGCAACTCTCGGGCGCACGT 29840
Qy 339 CCTGAGGCGCGGTGCGCGCGGTGCGAGTCCACCGGGTCTCGGAAGCGCTGGC 398
Db 29841 GCACGGGTGACGTGACCTGCTGCGGGGTGCGCCACGCGGCTCTGGTTCGACCTGCC 29900
Qy 399 AGCGAAGACGATCAGCAACTGCCACGCTGCTGCACACGATTCGCGCGCGGATCGC 458
Db 29901 CACCTACCTTCACAGGAGCGGCTCTGCGCCCAAGCCGACCGCAGGCGCACCTC 29960
Qy 459 GCGAAGAGGATCAGGCTCAACCGTCTCTTCAGCATGCTGCCCGCGCGAGCCGAA 518
Db 29961 GTCGCTGGGGTCCGTGAC-----TCGACCCACCGCTGTGACCGCGCAGT 30008
Qy 519 AGAGATGAAGTTCTTGAGCGACCCGAGATCGGTTCGGCTTATACAGCGCTTCGCGCGCA 578
Db 30009 CGACGTACCCGGTCAACGGCGAGCGGTGTTACCGGGCGGTCTCTCCCGACGAGCAGCA 30068
Qy 579 CTGGGACCGCTCGTCAATGCTGTGTGGGACCGGTCTGAGGTGGGTGAGGCGATCGG 638
Db 30069 CTGGCTACCCAGCACGCTGTGGTGGGCGCAACTGTCGCCGCGAGTTCCTGGTCA 30128
Qy 639 CCT---GCGCGCGCGGGTTCGACCTGCTCGCGCGCGCGCGCGGTGACCGTCTGCA 695
Db 30129 CTTGCGCTACCCCGCGGCGCGAGCTCGGCGTCCCGGTGTGGAGGAACTCGTCTGCA 30188
Qy 696 GCAGCTCCAGAGTGGCCAGCACGGAGAGTCTGTTTCCAGTCCGCGGAGACCGCGAA 755
Db 30189 GCAGCGGTGTTGTTGACCGCGCGGTGCTGCTGCGCTGCTGCGTTCGCGCGCGCGCA 30248
Qy 756 GG 757
Db 30249 CG 30250
```

RESULT 12
 US-09-252-991A-11386/c
 ; Sequence 11386, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 11386
 ; LENGTH: 534
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-11386

Query Match 4.3%; Score 50.2; DB 4; Length 534;
 Best Local Similarity 46.2%; Pred. No. 0.064;
 Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 447 CGCGGCGATCGCGCGAAACGGATCAGGCTCAACCCGTCCTTCGACGATGCTGCCCCG 506
 DB 463 CGCGGAGACCGGATCGGAGACCGCCAGGCTCACCGGAACTGGCGGTAGATGTTGCCAC 404

QY 507 GCGGAGCGAAGAGATGAAGTTCCTGAGCAGCCCGAGATCGTGGCTTATCAGCGC 566
 DB 403 CGCGCCGCTGAAGAACCCATCGGCACGAACACCGAGACGACCAACCGGTGATGCCGAC 344

QY 567 GCTTCGCGCCACTGGCGACCGCTCGTCATGCTGCTGGGACCGGTCCTGAGGTGGGG 626
 DB 343 GATGGCCCGCTGATCGCGCATCGCTTACCGTGGCGTCTGGCGGACAGGCTTC 284

QY 627 TGAGCGATCGGCTGGCGCGCGCGGTCGACCTGCTCGCCGCGCGCGCGCGCTGAC 686
 DB 283 CTCGGCCATCAGCGCTGCGACGTTCTCCACGAGATCGGTCGTCACCGAGATGCC 224

QY 687 CGTCGTCAGCAGCTCCAGAGCTGCCAGCGGAGAGCTCGTCTTCCAGTCCCGAA 746
 DB 223 GATCGCCAGGACCATCGGAACATGTCAGCAGTTGATCGAAGCCCGAGCGAT 164

QY 747 GACCGGAGGCGCGCGCGCGCGCTGAGTTTACACGAAAGTCGCTTACTGCTTACGC 805
 DB 163 CACCGTGAAGGTGCCAGCGGCGCCACCGGACCACTGTCGGGATCAGGCTGGCGC 105

RESULT 13
 US-09-252-991A-11487
 ; Sequence 11487, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 11487
 ; LENGTH: 3150
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-11487

Query Match 4.3%; Score 50.2; DB 4; Length 3150;
 Best Local Similarity 46.2%; Pred. No. 0.078;
 Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 447 CGCGGCGATCGCGCGAAACGGATCAGGCTCAACCCGTCCTTCGACGATGCTGCCCCG 506
 DB 1702 CGCGGAGACCGGATGAGAGACCGCGAGGTCACCGCAACTGGCGGTAGATGTTGCCAC 1761

QY 507 GCGGAGCGGAAAGAGATGAAGTTCCTGAGCAGCCCGAGATCGTGGCTTATCAGCGC 566
 DB 1762 CGCGCCGCTGAAGAACCGCATCGGCACGAACACCGAGACCGACCGGTGATGCCAC 1821

QY 567 GCTTCGCGCCACTGGCGACCGCTCGTCATGCTGCTGGCGACCGGTCGAGGTGGGG 626
 DB 1822 GATGGCCCGCTGATCTGGCGCATCGGCTTGACCGTGGCGTCTGTCGCGACAGGCTTC 1881

QY 627 TGAGCGATCGGCTGGCGCGCGCGGTCGACCTGCTCGCCGCGCGCGCGCGCTGAC 686
 DB 1882 CTCGGCCATCAGCCGCTCGACGTTCTCCACGATGATCGGTCGTCACGAGATGCC 1941

QY 687 CGTCGTCAGCAGCTCCAGGAGCTGGCCAGCAGGAGAGCTCGTCTTCCAGTCGCCGAA 746
 DB 1942 GATCGCCAGGACCATCGGAACATGTCAGCAGTTGATCGAAGCCCGAGCGAT 2001

QY 747 GACCGGAGGCGCGCGCGCGGTCAGTTTCACCGAAGTCTGCTTACTGCTTACGC 805
 DB 2002 CACCGTGAAGGTGCCAGCAGGCGCCACCGGACCAACGATGTCGGGATCAGGCTGGCGC 2060

RESULT 14
 US-09-252-991A-11405/c
 ; Sequence 11405, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 11405
 ; LENGTH: 3372
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-11405

Query Match 4.3%; Score 50.2; DB 4; Length 3372;
 Best Local Similarity 46.2%; Pred. No. 0.078;
 Matches 166; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 447 CGCGGCGATCGCGCGAAACGGATCAGGCTCAACCCGTCCTTCGACGATGCTGCCCCG 506
 DB 1458 CGCGGAGACCGGATGAGAGACCGCCAGGTCACCGGAACTGGCGGTAGATGTTGCCAC 1399

QY 507 GCGGAGCGGAAAGAGATGAAGTTCCTGAGCAGCCCGAGATCGTGGCTTATCAGCGC 566
 DB 1398 CGCGCCGCTGAAGAACCGCATCGGCACGAACACCGAGACCGACCGGTGATGCCGAC 1339

QY 567 GCTTCGCGCCACTGGCGACCGCTCGTCATGCTGCTGGCGACCGGTCGAGGTGGGG 626
 DB 1338 GATGGCCCGCTGATCTGGCGCATCGCTTGACCGTGGCGTCTGCGGACAGGCTTC 1279

QY 627 TGAGCGATCGGCTGGCGCGCGCGGTCGACCTGCTGCGCGCGCGCGCGCGCTGAC 686
 DB 1278 CTCGGCCATCAGCCGCTCGACGTTCTCCACGATGATCGGTCGTCACGAGATGCC 1219

QY 687 CGTCGTCAGCAGCTCCAGGAGCTGGCCAGCAGGAGAGCTCGTCTTCCAGTCGCCGAA 746

Db 1218 GATCCGAGGACCATGCCAATCGTCAGACGTTGATCGAGAGCCGCGGAGCAT 1159
Qy 747 GACCGGAGGCGCGGCGACCGTCAGTTTACACGAAAGTCGCTTACTGCTTACGC 805
Db 1158 CACCGTGAAGGTGCCGAGCGGCCACCGGACCACTGTCGGGATCAGGCTGGCGC 1100

RESULT 15
US-09-252-991A-2186
; Sequence 2186, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2186
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2186

Query Match 4.1%; Score 48.8; DB 4; Length 915;
Best Local Similarity 51.1%; Pred. No. 0.13;
Matches 140; Conservative 0; Mismatches 132; Indels 2; Gaps 1;
Qy 178 GCGATTTCTGGGGAGTGTGGCGGAGCTACGAAAGACGCTGAAACCGACCGCC--G 235
Db 395 GCCAGTTGCCGATGGACGTGGTGGCGGTGATTTCCAAATCACCAGCCTGGAGCGCTGG 454
Qy 236 TGAATCTCGAGGGGCAACCGGATCCGCAACCACTCTCTGCCATCTCGGCCATCTCACCC 295
Db 455 CGCGCTGGCAGCGCATTCCTACCACCACTTCCCGCTCGACCCGAAACGACAGCGCGGCC 514
Qy 296 TTGACGAGCTGACGGGAGGTCAACGACAGTGGGTCAACGACCTGGAGGCGCGGCTCG 355
Db 515 AGGAAGCCAGGGTCTGGCAGGTGCTCGAAGAGAGTGGCGCCGAGCTGGTGATCTCGGTC 574
Qy 356 GCCCGTGGCGGAGTCCACGCGGGTCTCGGAAGCGGCTGGCAGCGAAGACGATCAGCA 415
Db 575 GCTACATGAGGTGCTGTGCGCGAACTGTGCGGCGCTGGACGGCTGGGCGATCAACA 634
Qy 416 ATGCCACGGCTGTGTCACAGATCTCGGGGC 449
Db 635 TCCATCACTCCCTGTGCGCGGCTTCAGGGGCGC 668

Search completed: August 6, 2004, 13:42:24
Job time : 147.71 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: August 6, 2004, 08:50:04 ; Search time 1956.12 Seconds
(without alignments)
9439.175 Million cell updates/sec

Title: US-09-855-340A-2
Perfect score: 426
Sequence: 1 atcgcaacacacacggggct.....acctgttcgcggagctga 426

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.ov.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	426	100.0	426	6	AX338971 Sequence
2	426	100.0	2025	7	AY150027 Bacteriopl
3	66	15.5	1575	6	E03045 DNA encodin
4	64.6	15.2	1653	8	AK121255 Oryza sat
5	64.6	15.2	124629	8	OSU000082 Oryza sat
6	64.4	15.1	1056	6	BD269437 Mitomycin
7	64.4	15.1	1056	6	AR266923 Sequence
8	64.4	15.1	53500	6	BD269447 Mitomycin
9	64.4	15.1	53500	6	AR266933 Sequence
10	64.4	15.1	63734	1	AF127374 Streptomy
11	64.4	15.1	138203	1	AY310323 Streptomy
12	63.4	14.9	145539	8	CNS08CEV Oryza sat
13	62.6	14.7	276800	1	SC0939115 Streptomy
14	61.6	14.5	2081	1	SSBGL3GE Streptomyce
15	61.4	14.4	1233	6	BD179992 Highly th
16	61.2	14.4	10352	1	AF546152 Streptomy
17	61	14.3	110000	2	AC139251_1 Continuation (2 of
18	60.6	14.2	77457	1	AF210243 Streptomy
19	59.6	14.0	132544	1	AF521085 Streptomy
20	59	13.8	3198	1	AF233851 Streptomy
21	58.6	13.8	296300	1	AP005035 Streptomy
22	58.6	13.8	300425	1	AP005038 Streptomy
23	58.2	13.7	125020	9	AF429315 Home sapi
24	58.2	13.7	349315	1	BX572593 Rhodopseu
25	58	13.6	299423	1	AP005037 Streptomy
26	57.8	13.6	1193	8	AK109088 Oryza sat
27	57.8	13.6	143407	8	AP003825 Streptomy
28	57.8	13.6	299925	1	AP005045 Streptomy
29	57.6	13.5	120627	2	AC104714 Oryza sat
30	57.4	13.5	18209	1	AE005129 Halobacte
31	57.4	13.5	197050	1	AL646081 Ralstonia
32	57.4	13.5	321250	1	SC0939111 Streptomy
33	57.2	13.4	292100	1	SC0939121 Streptomy
34	57	13.4	15738	6	AX803762 Sequence
35	57	13.4	61944	6	AX803750 Sequence
36	56.8	13.3	972	6	AX574166 Sequence
37	56.8	13.3	1195	8	AK110593 Oryza sat
38	56.8	13.3	45055	6	AX574197 Sequence
39	56.8	13.3	138390	14	AY261359 Bovine he
40	56.6	13.3	303550	1	SC0939131 Streptomy
41	56.4	13.2	1381	14	AY299389 Ori virus
42	56.4	13.2	58343	1	AB089954 Micromono
43	56.4	13.2	310550	1	SC0939113 Streptomy
44	56.2	13.2	420	1	MXU49152 Myxococcus
45	56.2	13.2	923	8	AK063822 Oryza sat

ALIGNMENTS

RESULT 1
AX338971
LOCUS AX338971
DEFINITION Sequence 2 from Patent WO0187936.
ACCESSION AX338971
VERSION AX338971.1 GI:18129107
KEYWORDS
SOURCE Micromonospora carbonacea
ORGANISM Micromonospora carbonacea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE
1
Hosted, T.J. and Horan, A.C.
Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into

Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 22-NOV-2001;
SCHERING CORPORATION (US)

FEATURES

source
1. .426
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"

ORIGIN

Query Match 100.0%; Score 426; DB 6; Length 426;
Best Local Similarity 100.0%; Pred. No. 5.4e-42;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGMAACACACGGGGCTGGGGCGGGGACATATGGGGCGGCATACCTCTACCGCCCGC 60
DB 1 ATGGCGAACACACGGGGCTGGGGCGGGGACATATGGGGCGGCATACCTCTACCGCCCGC 60

QY 61 GAGCGCGCGGACTGACCAAGACGAGTTGGCCAGCGGCATCCAGAGGACCGGGCCACC 120
DB 61 GAGCGCGCGGACTGACCAAGACGAGTTGGCCAGCGGCATCCAGAGGACCGGGCCACC 120

QY 121 GTGGGCGGTGGAGGACGCAAGAACCGGCCCGGACGCGGCATCTGTTGCCGCGTC 180
DB 121 GTGGGCGGTGGAGGACGCAAGAACCGGCCCGGACGCGGCATCTGTTGCCGCGTC 180

QY 181 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGCATATCGCCCTATCTGGAGCGCGT 360
DB 181 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGCATATCGCCCTATCTGGAGCGCGT 360

QY 241 GTCAACCGCGGACGACCGCCCAACCATGACCTGGACCTGGAGGAAATCGAGTGTGCGCAC 300
DB 241 GTCAACCGCGGACGACCGCCCAACCATGACCTGGAGGAAATCGAGTGTGCGCAC 300

QY 301 GACCCCAAGTGCAGGAGCATGACGCGCGCATATCGCCCTATCTGGAGCGCGT 360
DB 301 GACCCCAAGTGCAGGAGCATGACGCGCGCATATCGCCCTATCTGGAGCGCGT 360

QY 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACCTGTTCCGCGG 420
DB 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTATCGACCTGTTCCGCGG 420

QY 421 AGCTGA 426
DB 421 AGCTGA 426

RESULT 2

AY150027
LOCUS Bacteriophage pMLP1 att/int region.
DEFINITION Bacteriophage pMLP1 att/int region.
ACCESSION AY150027
VERSION AY150027.1 GI:28630421

KEYWORDS

Source

ORGANISM

Viruses.

1 (bases 1 to 2025)

Alexander, D.C., Devlin, D.J., Hewitt, D.D., Horan, A.C. and

Hosted, T.J.

Development of the Micromonospora carbonacea var. africana ATCC

39149 Bacteriophage pMLP1 integrase for site-specific integration

in Micromonospora spp

Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)

12949170

2 (bases 1 to 2025)

Hosted, T.J. Jr., Alexander, D.C. and Hewitt, D.D.

Direct Submission

Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough

Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA

Location/Qualifiers

1. .2025

/organism="Bacteriophage pMLP1"

/mol_type="genomic DNA"
/specific_host="Micromonospora sp. ATCC 39149"
/db_xref="taxon:21922"
/note="host is deposited in ATCC as Micromonospora carbonacea var. africana Waltz et al."
1. .2025
/note="att/int region"
197. .622
/note="xis"
/codon_start=1
/transl_table=11
/product="excisionase"
/protein_id="AAO46044.1"
/db_xref="GI:28630422"
/translation="MRNTPGLRGTTWAAAYVLTARERAGLTYSLARIRIQKQRTAVGRW
EDGNRPDADLVARVAQVLGLDDEAAGAARPGVTPATPTMDLDEIELVRIDP
KLDEMKRRIIALILERRERDKAAATSETRKLDLFRS"
628. .1806
/gene="intM"
628. .1806
/gene="intM"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="AAO46045.1"
/db_xref="GI:28630423"
/translation="MWIEKNGPVVIRIDLVRGKVTIOTGYPTKSAKNAMVOFRAEQ
LOGNALMPRGGOITLADPVGEMWPSYEKTLKPTAVNSEGNRIHLLPILGHITLDEL
DQVQTVWVNDLEAGVGPWEESTGRKPLAAKTI SNCHGLLHTI GGAIAAKRIRLN
PCSTNLPFRPEKMFSLDPEIGRLITLAPPHMRPLVLLVAIGKRWAGIAIGRAGR
VDLAAAPKLVVEQELASTGELVFQSPKAGRTVSTFTKVALLLTLPLAGKKS
DEVVTAAPKGMVTRFRIRIVKACEEAGLPGLRIHLRHTAAI.II SAGRPLSAIS
RLHSHSIAVTDLILYGLHREVDGILAAIEEAMAGVRAEDLEAEDELDTDLADAA
"
join(1984. .1895,1900. .1911)
stem loop
misc_feature
1925. .1986
/note="attP element"
join(1969. .1986,1991. .2010)
ORIGIN

Query Match 100.0%; Score 426; DB 7; Length 2025;
Best Local Similarity 100.0%; Pred. No. 3.5e-42;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCAACACACCGGGCTGGGGCGGGGACATATGGGGCGGCATACGTCCTACCGCCCGC 60
DB 197 ATGCGCAACACACCGGGCTGGGGCGGGGACATATGGGGCGGCATACGTCCTACCGCCCGC 256

QY 61 GAGCGCGCGGACTGACCAAGACGAGTTGGCCAGCGGCATCCAGAGGACCGGGCCACC 120
DB 257 GAGCGCGCGGACTGACCAAGACGAGTTGGCCAGCGGCATCCAGAGGACCGGGCCACC 316

QY 121 GTGGCGCGGTGGGAGGACGCGCAAGAACCGGGCGGACGACGCGGACCTCGTGTCCCGGTC 180
DB 317 GTGGCGCGGTGGGAGGACGCGCAAGAACCGGGCGGACGACGCGGACCTCGTGTCCCGGTC 376

QY 181 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGCATCTGCGCCCGCGG 240
DB 377 GCCAGGTGTCGGCCTCGACCTCGACGAGCCCTCGCGCGCGCATCTGCGCCCGCGG 436

QY 241 GTCAACCGCGGACGACCGCCCAACCATGACCTGGACGAGGAAATCGAGCTGTGTCGCGCAC 300
DB 437 GTCAACCGCGGACGACCGCCCAACCATGACCTGGACGAGGAAATCGAGCTGTGTCGCGCAC 496

QY 301 GACCCCAAGTGCAGGAGCATGACGCGCGGCATATCGCCCTATCTGGAGCGCGCT 360
DB 497 GACCCCAAGTGCAGGAGCATGACGCGCGGCATATCGCCCTATCTGGAGCGCGCT 556

QY 361 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCGG 420
DB 557 GAGCGGACAAAGCGGGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCGG 616

QY 421 AGCTGA 426

```

Db      617 AGCTGA 622      |||||
E03045 1575 bp      DNA      linear      PAT 29-SEP-1997
LOCUS   DNA encoding streptomycetes chromofuscns phosphatidyl choline
DEFINITION
E03045  phosphatidohydrolase.
VERSION  E03045.1  GI:2171267
KEYWORDS Streptomycetes chromofuscus
SOURCE   JP 1991187382-A/1.
ORGANISM Streptomycetes chromofuscus
REFERENCE
AUTHORS  Houriyouchi,K., Mizoguchi,J., Takahara,M., Imamura,S., Beppu,T. and
          Horinouchi,S.
TITLE    DNA CONTAINING GENETIC INFORMATION OF PHOSPHOLIPASE AND USE THEREOF
JOURNAL  Patent: JP 1991187382-A 1 15-AUG-1991;
COMMENT  TOYO JOZO CO LTD
OS       Streptomycetes chromofuscus
PN       JP 1991187382-A/1
PD       15-AUG-1991
PF       15-DEC-1989 JP 1989325355
PI       HOURIYOU KAZDO, MIZOGUCHI JUNZO, TAKAHARA MASAYASU, PI
          IMAMURA SHIGEVUKI,
          BEPPU TERUHIKO, HORINOUCI SUEJI
PC       C12N15/55,C12N1/19,C12N1/21,C12N9/20, (C12N1/19,C12R1:865), PC
          (C12N1/21,
          C12R1:19), (C12N1/21,C12R1:465), (C12N1/21,C12R1:125); CC
          strandedness: Double;
          CC topology: Linear;
          CC hypothetical: No;
          CC anti-sense: No;
          CC *source: strain=A-0848;
          CC *source: developmental_stage=Germ-Line;
          CC *source: clone=pcPLDI;
          PH Key      Location/Qualifiers
          FT mat_peptide 43..1572
          FT /product="Streptomycetes chromofuscus FT
          FT Phosphatidylcholine
          FT Phosphatidohydrolase'.
          FT Location/Qualifiers
          FT 1..1575
          FT /organism="Streptomycetes chromofuscus"
          FT /mol_type="genomic DNA"
          FT /db_xref="taxon:42881"

FEATURES
Source
Location/Qualifiers
1..1575
/organism="Streptomycetes chromofuscus"
/mol_type="genomic DNA"
/db_xref="taxon:42881"

ORIGIN
Query Match 15.5%; Score 66; DB 6; Length 1575;
Best Local Similarity 48.4%; Pred. No. 29;
Matches 183; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

QY 7 AACACACCGGGGTGGGGCGGCGACATGGCGCGATAGTCTCTACCGCCCGCGAGCGC 66
|||
Db 646 AAGCGCCGGTCTCCGATCTGGGACGACACGAGTTCCGACACACCCCTTGTCGGCG 705
|||
QY 67 GCGGACTGACCAAGAGCGAGTTGGCAGCGCGCATCCAGAAGGACCGGGCCACCGTCGGC 126
|||
Db 706 GCGCGGTGAACACACCGAGGCGCGCGGACCTGTCGGCGCGTCAGGCGCGCGCC 765
|||
QY 127 CGTGGGAGGACGGCAGAACCGGCCGAGCGGAGCTGTTGCCGCGTCGCCGCG 186
|||
Db 766 AAGCAGGCTTACTTCGAGTGGATGCGCGGCGCGCCCATCGCGGACCACTACCGG 825
|||
QY 187 GTCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGAGGTCGTGCGCCCGCGCTACCC 246
|||
Db 826 CGGCTCGGCTTCGGCAAGCTCGCGGACCTCTCCTCTGACCTGCGCTCTCTCGGCTCC 885
|||
QY 247 CGGCCAGCGACCCCAACCACTGAGCCTGGAGAGGAAATCGAGCTGGTCCGCAACGCC 306
|||

```

```

Db      886 CAGCAGGCTCCACGGCCAGCGGTTGCTGGAGACCCGCGACCGTACGCTCACCGCGC 945
|||
QY 307 AAGCTGGACGAGGACATGAGCGCGGATCATGCCCTATCTCTGGAGCGCGCTGACGCGC 366
|||
Db 946 GCGAGCTCGACTCGCTCAAGGGCGGCTGAAGCGCTCGGACACCGAGGTGGCGCTGGTC 1005
|||
QY 367 GACAAGGCGCGCGGCGATC 384
|||
Db 1006 GGCACACTCCGTGATGATC 1023
|||

RESULT 4
AKI21255 1653 bp      mRNA      linear      PLN 29-OCT-2003
LOCUS   Oryza sativa (japonica cultivar-group) cDNA clone:J023102B22, full
DEFINITION
          insert sequence.
ACCESSION AKI21255.1 GI:37990878
VERSION   AKI21255
KEYWORDS  FLI_CDNA; CAP trapper.
SOURCE    Oryza sativa (japonica cultivar-group)
ORGANISM  Oryza sativa (japonica cultivar-group)
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
1 The Rice Full-length cDNA Consortium, National Institute of
  Agrobiological Sciences Rice Full-length cDNA Project Team,
  Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
  Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
  Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
  Ohsuki,K., Shishiki,T., Foundation of Advancement of International
  Science Genome Sequencing & Analysis Group; Otonoda,Y.,
  Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
  Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
  Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
  Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
  Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN;
  Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
  Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
  Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
  Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
  Yoshino,M. and Hayashizaki,Y.
  Collection, mapping, and annotation of over 28,000 cDNA clones from
  japonica rice
  Science 301 (5631), 376-379 (2003)
  227552273
  12869764.

2
  Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
  Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
  Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
  Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,K., Imamura,K.,
  Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
  Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
  Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
  Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
  Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
  Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
  Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakahama,I.,
  Nakamura,M., Namiki,T., Narikawa,R., Niikura,J., Nisui,K.,
  Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M.,
  Ooka,H., Osato,N., Ota,Y., Otonoda,Y., Ryo,R., Saitoh,H., Sakai,C.,
  Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K.,
  Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y.,
  Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M.,
  Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
  Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K.,
  Xie,Q., Yahegi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J.,
  Yokomizo,S. and Yoshimura,A.
  Collection, mapping, and annotation of 28K full-length cDNA clones
  from japonica rice
  Unpublished

```



```

Db 46991 GGCTGCGCGAGGCGCGCGCGCTCTCGGAGGACGCCGAGCGCCGTACCCCGGTG 47050
QY 271 CTGGACGAGGAATCGAGCTGTGTCGCCACCGACCCCAAGCTGGACGAGACATGAAGCGG 330
Db 47051 GTGGGAGAGATGATGAAGCGCGCGCGGCGGCGGCGGAGCGCGTGTCTATGG 47110
QY 331 CGCATCATCGCCCTAATCTCTGAGCGCCGTGAGCGGCAAGCGGCGCGCATCGAGGAA 390
Db 47111 GCGGTCTGCCACAAGTACAGGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 47170
QY 391 ACCAAGCGCTCATCGACCTGTCTTCGCGCGGAGCTG 425
Db 47171 GGCTGACTCGCCTCTCTCTGCTGTCGAGAGCGG 47205

RESULT 6
LOCUS BD269437 1056 bp DNA linear PAT 17-JUL-2003
DEFINITION Mitomycin biosynthetic gene cluster.
ACCESSION BD269437
VERSION JP 2002537833-A/51.
KEYWORDS Streptomyces lavendulae
SOURCE Streptomyces lavendulae
ORGANISM Streptomyces lavendulae
REFERENCE 1 (bases 1 to 1056)
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.C.
TITLE Mitomycin biosynthetic gene cluster
JOURNAL Patent: JP 2002537833-A 51 12-NOV-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA
COMMENT OS Streptomyces lavendulae
PN JP 2002537833-A/51
PD 12-NOV-2002
PF 10-MAR-2000 JP 2000603359
PR 12-MAR-1999 US 09/266965
PI DAVID H SHERMAN, YINGQING MAO, MUSTAFA VAROGLU, MIN HE, PAUL C PI
SHELDON
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P17/10 PC
C12O1/68, G01N33/53,
PC G01N33/566//C12N9/00, (C12P17/10, C12R1:465), C12N15/00, C12N5/00
CC Mitomycin biosynthetic gene cluster
FH Key Location/Qualifiers
FT source 1..1056
FT /organism='Streptomyces lavendulae'.
FEATURES
source Location/Qualifiers
1..1056
/organism="Streptomyces lavendulae"
/mol_type="genomic DNA"
/db_xref="taxon:1914"

ORIGIN
Query Match 15.1%; Score 64.4; DB 6; Length 1056;
Best Local Similarity 51.0%; Pred. No. 50;
Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAGACCGCGGCCACCGTCCGCGCGGTGGAGGACGCGAAGACCGCG 151
Db 683 CCGGCGCGGCTCTCGCGGACCGGTGGAGATCTCTGCGCGGACCTCTTCGAGACCATCC 742
QY 152 CCGACGACGCGGACCTCTGTTCCCGCGGTCCCGCAGGTCTCGGCTCGACCTCGAGGAG 211
Db 743 CCGACGCGCGGACGCTTACTCTCATAGCACGTGTCTGACGACTGGAGACGACGACG 802
QY 212 CCTCTCGCGCGGACGCTCTCGCGCGCGGTACCCCGCAGGACCCCAACCATGGACC 271
Db 803 TCGTACGCATCTCTCGCGGACCGGTCCACCGCATGAAGCGGACTCCCGGCTCTGTGTA 862
QY 272 TGGACGAGGAATCGAGCTGTGTCGCCACCGACCCCAAGCTGGAGGACATGAAGCGGC 331
Db 863 TCGACAACTCATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTCTGCTGC 922
QY 332 GCATCATCGCCCTAATCTCTGGAGCGCGGTGAGCGGCAAGCGCGGCGCATCGAGGA 389
Db 923 TCGTCTCTCTCGCGGCGCGGACCGTTCGAGAGCGAATTCGCGCGCTCTCGAGAA 980

RESULT 8
LOCUS BD269447 53500 bp DNA linear PAT 17-JUL-2003
DEFINITION Mitomycin biosynthetic gene cluster.
ACCESSION BD269447
VERSION BD269447.1 GI:33079215
KEYWORDS JP 2002537833-A/61.
SOURCE Streptomyces lavendulae
ORGANISM Streptomyces lavendulae
REFERENCE 1 (bases 1 to 53500)
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.C.
TITLE Mitomycin biosynthetic gene cluster
JOURNAL Patent: JP 2002537833-A 61 12-NOV-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA
COMMENT OS Streptomyces lavendulae
PN JP 2002537833-A/61
PD 12-NOV-2002
PF 10-MAR-2000 JP 2000603359
PR 12-MAR-1999 US 09/266965
PI DAVID H SHERMAN, YINGQING MAO, MUSTAFA VAROGLU, MIN HE, PAUL C PI
SHELDON
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P17/10 PC

```

```

Db 923 TCGTCTCTCTCGCGGCGCGGACCGTTCGAGAGCGAATTCGCGCGCTCTCGAGAA 980

RESULT 7
LOCUS AR266923 1056 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 66 from patent US 6495348.
ACCESSION AR266923
VERSION AR266923.1 GI:29696305
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1056)
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.
TITLE Mitomycin biosynthetic gene cluster
JOURNAL Patent: US 6495348-A 66 17-DEC-2002;
FEATURES Location/Qualifiers
source 1..1056
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 15.1%; Score 64.4; DB 6; Length 1056;
Best Local Similarity 51.0%; Pred. No. 50;
Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAGACCGCGGCCACCGTCCGCGCGGTGGAGGACGCGAAGACCGGC 151
Db 683 CCGGCGCGGCTCTCGCGGACCGGTGGAGATCTCTGCGCGGACCTCTTCGAGACCATCC 742
QY 152 CCGACGACGCGGACCTCTGTTCCCGCGGTCCCGCAGGTCTCGGCTCGACCTCGAGGAG 211
Db 743 CCGACGCGCGGACGCTTACTCTCATAGCACGTGTCTGACGACTGGAGACGACGACG 802
QY 212 CCTCTCGCGCGGACGCTCTCGCGCGCGGTACCCCGCAGGACCCCAACCATGGACC 271
Db 803 TCGTACGCATCTCTCGCGGACCGGTCCACCGCATGAAGCGGACTCCCGGCTCTGTGTA 862
QY 272 TGGACGAGGAATCGAGCTGTGTCGCCACCGACCCCAAGCTGGAGGACATGAAGCGGC 331
Db 863 TCGACAACTCATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTCTGCTGC 922
QY 332 GCATCATCGCCCTAATCTCTGGAGCGCGGTGAGCGGCAAGCGCGGCGCATCGAGGA 389
Db 923 TCGTCTCTCTCGCGGCGCGGACCGTTCGAGAGCGAATTCGCGCGCTCTCGAGAA 980

RESULT 8
LOCUS BD269447 53500 bp DNA linear PAT 17-JUL-2003
DEFINITION Mitomycin biosynthetic gene cluster.
ACCESSION BD269447
VERSION BD269447.1 GI:33079215
KEYWORDS JP 2002537833-A/61.
SOURCE Streptomyces lavendulae
ORGANISM Streptomyces lavendulae
REFERENCE 1 (bases 1 to 53500)
AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.C.
TITLE Mitomycin biosynthetic gene cluster
JOURNAL Patent: JP 2002537833-A 61 12-NOV-2002;
REGENTS OF THE UNIVERSITY OF MINNESOTA
COMMENT OS Streptomyces lavendulae
PN JP 2002537833-A/61
PD 12-NOV-2002
PF 10-MAR-2000 JP 2000603359
PR 12-MAR-1999 US 09/266965
PI DAVID H SHERMAN, YINGQING MAO, MUSTAFA VAROGLU, MIN HE, PAUL C PI
SHELDON
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P17/10 PC

```



```
.C12Q1/68.G01N33/53,
PC GOIN33/566/C12N9/00,(C12P17/10.C12R1.465),C12N15/00,C12N5/00
CC Mitomycin biosynthetic gene cluster
FH Key Location/Qualifiers
FT source 1..53500
FT /organism='Streptomyces lavendulae'.
FEATURES
    source
        Location/Qualifiers
            1..53500
            /organism='Streptomyces lavendulae'
            /mol_type='genomic DNA'
            /db_xref='taxon:1914'
ORIGIN
    Query Match 15.1%; Score 64.4; DB 6; Length 53500;
    Best Local Similarity 51.0%; Pred. No. 17;
    Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 92 CCAGCGCGCATCCAGAAAGACCGGCGCCACCGCTCGCGCGGTGGGAGACCGCAAGACCGGC 151
DB 43516 CCGGCGCGGCGCTCGCGGACCGGTGCGAGATCTTCCGCCGCGACTTCTTCGAGACCATCC 43575
QY 152 CCGAGCAGCGGACCTCGTGTGCCCGCTGCGCCAGTCTCGGCTCGACCTCGACGAG 211
DB 43576 CCGAGCGGCGCGACGTCTACCTCATCAAGCAGCTGCTGCAGCACTGGGACGACGACG 43635
QY 212 CCTCGCGCGCGACGCTGTGCCCGCGCGGTCAACCGCGGACCGACCCCAACCATGGACC 271
DB 43636 TCGTACGACATCTCGCGCGGATCGCCACCGGATGAAGCGGACTCCCGGCTCTGTGCA 43695
QY 272 TGGACGAGGAATCGAGTGTGTCGACCGACCCCAAGCTGAGCAGGACATGAAGCGGC 331
DB 43696 TCGAACAACCTCATCGACGAGCGGCGCGCGCATCGACGCTCTTGTGACCTGTCTGTC 43755
QY 332 GCATCATCGCCCTAATCTCTGAGCGCCGTGAGCGGACAAAGCGCGCGCATCGAGGA 389
DB 43756 TCGTCTCTGTGGCGGCGGACCTCGGAGAGCGAATTCGCGCGCTGCTGAGAA 43813
RESULT 9
LOCUS AR266933 53500 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 76 from patent US 6495348.
ACCESSION AR266933
VERSION AR266933.1 GI:29696315
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
    1 (bases 1 to 53500)
    AUTHORS Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.
    TITLE Mitomycin biosynthetic gene cluster
    JOURNAL Patent: US 6495348-A 76 17-DEC-2002;
    LOCATION/Qualifiers
    source
        1..53500
        /organism='unknown'
        /mol_type='genomic DNA'
ORIGIN
    Query Match 15.1%; Score 64.4; DB 6; Length 53500;
    Best Local Similarity 51.0%; Pred. No. 17;
    Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 92 CCAGCGCGATCCAGAAAGACCGGCGCCACCGCTCGCGCGGTGGGAGACCGCAAGACCGGC 151
DB 43516 CCGGCGCGGCGCTCGCGGACCGGTGCGAGATCTTCCGCCGCGACTTCTTCGAGACCATCC 43575
QY 152 CCGAGCAGCGGACCTCGTGTGCCCGCTGCGCCAGTCTCGGCTCGACCTCGACGAG 211
DB 43576 CCGAGCGGCGGACGTCTACCTCATCAAGCAGCTGCTGCAGCACTGGGACGACGACG 43635
QY 212 CCTCGCGCGCGACGCTGTGCCCGCGCGGTCAACCGCGGACCGACCCCAACCATGGACC 271
DB 43636 TCGTACGACATCTCGCGCGGATCGCCACCGGATGAAGCGGACTCCCGGCTCTGTGCA 43695
```

```
QY 272 TGGACGAGGAATCGAGTGTGTCGCGACCGACCCCAAGCTGGAAGGACATGAAGCGGC 331
DB 43696 TCGAACAACCTCATCGAGCAGCGGCGCGCGCATCGAGCTCTTGTGACCTGTCTGTC 43755
QY 332 GCATCATCGCCCTAATCTCTGAGCGCGGTGAGCGGACAAAGCGCGCGCATCGAGGA 389
DB 43756 TCGTCTCTGTGGCGGCGGACCGTTCGGAGAGCGAATTCGCGCGCTGCTGAGAA 43813
RESULT 10
LOCUS AF127374 63734 bp DNA linear BCT 27-MAY-1999
DEFINITION Streptomyces lavendulae LinA homolog, cytochrome P450 hydroxylase
ORF4, cytochrome P450 hydroxylase ORF3, MitT (mitT), MitS (mitS),
MitR (mitR), MitQ (mitQ), MitP (mitP), MitO (mitO), MitN (mitN),
MitM (mitM), MitL (mitL), MitK (mitK), MitJ (mitJ), MitI (mitI),
MitH (mitH), MitG (mitG), MitF (mitF), MitE (mitE), MitD (mitD),
MitC (mitC), MitB (mitB), MitA (mitA), MmCA (mmCA), MmCB (mmCB),
MmCC (mmCC), MmCD (mmCD), MmCE (mmCE), MmCF (mmCF), MmCG (mmCG),
MmCH (mmCH), MmCI (mmCI), MmCJ (mmCJ), MmCK (mmCK), MmCL (mmCL),
MmCM (mmCM), MmCN (mmCN), MmCO (mmCO), MmD (mrd), MmCP (mmCP), MmCQ
(mmCQ), MmCR (mmCR), MmCS (mmCS), MmCT (mmCT), MmCU (mmCU), MmCV
(mmCV), MmCW (mmCW), MmCX (mmCX), and MmCY (mmCY) genes,
complete cds, and unknown genes.
ACCESSION AF127374.1 GI:4731328
VERSION AF127374
KEYWORDS Streptomyces lavendulae
SOURCE Streptomyces lavendulae
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces; Streptomycetaceae; Streptomyces.
REFERENCE
    1 (bases 1 to 63734)
    AUTHORS Mao,Y., Varoglu,M. and Sherman,D.H.
    TITLE Molecular characterization and analysis of the biosynthetic gene
    cluster for the antitumor antibiotic mitomycin C from Streptomyces
    lavendulae NRRL 2564
    JOURNAL Chem. Biol. 6 (4), 251-263 (1999)
    MEDLINE 99201491
    PUBMED 10099135
REFERENCE
    2 (bases 1 to 63734)
    AUTHORS Mao,Y.Q., Varoglu,M. and Sherman,D.H.
    TITLE Direct Submission
    JOURNAL Submitted (10-FEB-1999) Microbiology, Uni. of Minnesota, Box196,
    420 Delaware St. SE., Minneapolis, MN 55455, USA
REFERENCE
    3 (bases 1 to 63734)
    AUTHORS Mao,Y.Q., Varoglu,M. and Sherman,D.H.
    TITLE Direct Submission
    JOURNAL Submitted (27-MAY-1999) Microbiology, Uni. of Minnesota, Box196,
    420 Delaware St. SE., Minneapolis, MN 55455, USA
REMARK Amino acid sequence updated by submitter
FEATURES
    Location/Qualifiers
        1..63734
        /organism='Streptomyces lavendulae'
        /mol_type='genomic DNA'
        /strain='NRRL 2564'
        /db_xref='taxon:1914'
        123..1367
        /function='dehydrogenase'
        /notes='ORF6'
        /codon_start=1
        /transl_table=11
        /product='unknown'
        /protein_id='AAD28446.1'
        /gb_xref='GI:4731329'
        /translation='MRETLFELGDIPIPTVPEQMYASVIRERYGQHPQAFREVDV
        VPKVGPQALVLMAGINYNVWASLGQPDVVISARQKHSEDFHIGSESGVYV
        AVDEGTVQAVGDEVLTSGQWTEADIRLGAADPTSGSQSVNGYEGYSPAQFAL
        VDDYQCHPKPGLTWEEAACFLTGTATYRLCGWPHDVRPGDPLVIGWAGLGISM
        AIOITRAGGIPVAVYDDEERASYCRLGAQGTINRLDFDHGRLPLDIGHAMGRWT
        EGYRAGRPWEVLGERSPRIVLEHSGOATPISMYLCDNAGMVIICGTTGYNADI
        DLRFLLMRQRLQGSFANRLQRCDVILHWANGQLDPCLSWTGGFDIGAKQLMHDN
        QHPQGNQAVLVNAPRTGLTFFA'
        CDS
```



```

Db      50316  CCGGCGCGCGCTCGCGACCGGTGGGAGATCCTCGCGCGGACTTCTTCGAGACCATCC 50375
Qy      152    CCGAGCAGCGGACCTCGTTCGCCCGCTGCGCCAGGTGCTCGGCTCGACCTCGACGAAG 211
Db      50376  CCGAGCGCGCGCGCTACCTCATCAGCACGTGCTGCAGACTGGGACGACGACG 50435
Qy      212    CCTCGCGCGCGAGGTCTGCGCGCGCGGTCAACCCCGCCAGCGACCCACCATGGACC 271
Db      50436  TCGTAGCATCTCCGCGCGGATCGCCACCGCATGAAGCGGACTCCCGGCTCTGGTCA 50495
Qy      272    TCGACGAGAAATCAGAGTGTGTCGCACCGACCCCAAGCTGCAGCAGACATGAAGCGGC 331
Db      50496  TCGAACACCTCATCAGCAGCGCGCGCGCATCGAGCTCTTCGTCGACCTGCTGCTGC 50555
Qy      332    GCATCATCGCCTATCTCGAGCGCGGTGACGCGGACAAAGCGCGCGCATCGAGGA 389
Db      50556  TCGTCTCTGTCGCGCGCGCAACGCTCGAGAGCGAATTCGCCCGCTGCTGGAGAA 50613

RESULT 11
LOCUS   AY310323/c 138203 bp DNA linear BCT 06-DEC-2003
DEFINITION Streptomyces sp. FR-008 heptaene macrolide complex synthesis gene
cluster, complete sequence.
ACCESSION AY310323
VERSION   AY310323.1 GI:34766435
KEYWORDS
SOURCE   Streptomyces sp. FR-008
ORGANISM Streptomyces sp. FR-008
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 138203)
AUTHORS  Chen,S., Huang,X., Zhou,X., Bai,L., He,J., Jeong,K.J., Lee,S.Y. and
Deng,Z.
TITLE    Organizational and Mutational Analysis of a Complete
FR-008/Candidin Gene Cluster Encoding a Structurally Related
Polyene Complex
JOURNAL  Chem. Biol. 10 (11), 1065-1076 (2003)
PUBMED  14652074
REFERENCE 2 (bases 1 to 138203)
AUTHORS  Chen,S., Huang,X., Zhou,X., He,J., Bai,L., Jeong,K.J., Lee,S.Y. and
Deng,Z.
TITLE    Direct Submission
JOURNAL  Submitted (30-MAY-2003) Bio-X Life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China

FEATURES
source   1..138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
complement(574..1950)
/gene="fscO"
complement(574..1950)
/gene="fscO"
/notes="putative FAD-dependent monoxygenase"
/codon_start=1
/transl_table=11
/product="fscO"
/protein_id="AAQ82549.1"
/db_xref="GI:34766436"
/translation="WPSTCPAPTSIGFRGCPHRKGRTPGRRITAVTPRCQKVRKH
QAKKRVSMRGSVAVVGGSTAGCAALASRGGAERTVFERADARLDRGVGIGV
ELRRVPATYRAGAKVEAVEQDADGTVRLADGEPRFDLVGADGYRNVREAMY
PGTGAAYAGTIGWRTSDVSLPSDGLDAHNTFPFGCHCWAYRIPDGSGGHRLNWVL
YTPAPRIDGLDLPTPTSLPPGRNLASLTELRLVAEHPFPFWAAKLSTPAETTF
IOPYDLDPVHVATDRMWLIGDAASVAREPHLGASVKALQDATALAEAWAGESWKEI
LEGYHAAGPVGVTAMVLGRMGSQAQVEZTPDWSAMGQAEFDANWQEQNGSDRSRSG
GGHSLKSR"
complement(2264..3037)
/gene="pabC"
complement(2264..3037)

```

```

/gene="pabC"
/notes="putative 4-amino-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="pabC"
/protein_id="AAQ82550.1"
/db_xref="GI:34766437"
/translation="MIEDGEPAGPEALASLASLTNYGHFTTLLVNGRVRGLDLHLER
LIRDTLFDALDPRVKLARAAPTGRATRVTVDFALNLGNIAADAPGILV
TSRPADKPFGLPRVSRVHRRDLPEKSVGLCTPLRLRQAQAGYDDVLYFTGPDG
ILGEGTNVGLVDGVEVWPGGVLAGTTQLLRGRHRSRHTCRPRGPRLSRGLRD
ORRVGVRPV"GIIDREFPAAHSHVTRLAETIYQALPGSPL"
complement(3150..3818)
/gene="fscRI"
complement(3150..3818)
/gene="fscRI"
/notes="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="fscRI"
/protein_id="AAQ82551.1"
/db_xref="GI:34766438"
/translation="MDPAPAADPAGSPERAANVLADGFDAGAYVACLDPSLTITQQNVQ
EFERRFGSSSELCSGRCFCDLVHPVQQQLMHQFARMLDGKRHRFATEVIAVQBERTA
STPLNALAVRGRTPDVAAILVVMNAEEEGDADVMAPRKLLSEIDARILEGIAA
GVSTIPLASRLYLSRQGVVHTGLLRLKLVNRAALVSRAYSMGVLKVGTPWPKVVE
DFIK"
complement(4377..7205)
/gene="fscRII"
complement(4377..7205)
/gene="fscRII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRII"
/protein_id="AAQ82552.1"
/db_xref="GI:34766439"
/translation="MLIERTTEIALVDAALGAATEGRSSVLITGLIGTSALLQRL
SCLADDRDVRVLRHAAPMEQDFAGVVRQLFETLLIGDSPEDARWMDAHASFARH
VLADDAAPPQADQAIATBAVLHGLLSLANVSADSKLLILVDDLOWSDVPVSRWLTY
LAKRLHGLRAVVCALRDGDPDRSHHTLAREIAREAGTQTLRPASLTATHELVRHFG
EAGDDFVQCHESAVGNPLFLLSILVGTGFLGRRPLAHEATARILRPSQLERLAS
ILRTQAPVRDLAAAIILGEGSDAPTLARLADGDSIGYAGALRAGALGALAAAPDEP
RFTHRSVRDAAESTLWQREVRHDEAAALLYAAGSPAQVAQALMAVVTTRQFVAD
VLRAAADTALRKGAPDTAAGYLCRALLDSFPAAGVGRGLLVELGTAEGRDFLACERH
IAQAMTLLPPDRVAALRISPTALGPAPLTAVDLLRQAAEDLGPADGLTGTDRDLA
LRLEARLHCGHEDPHELAASARHLKELGPEPVDVTGGRELTIALLGAATIGCHLPA
SEVVRLAGVLDREPATSRIHSTLPLAITMTAADSVAEVDSMLAVEHANEGRGTP
ANALVIEQALVHLGRGLAQAREQSETALGMTEPAHQGLDTATHTHALVALESDDP
SLAEKVLRRASRIRPQGLTTLGILRLLEATREIAAGDQTCALDSLSCGRLQEAAGWH
NPVLPFRPWAVGVHRRLDGDIRTARSLAEFEYAEQWGAQVGVGRALRLLEDGGH
RGIGLREAVTVLRGSANQLARALSTLAERLGGGAETEALAREATLATSQCAPGL
AERARAGAGHRSGPPATPEDVLTPTERRVVVGLACRGLTNGEIAAGALGVSSRAVEKHIT
HAVRULGISGRRELI VLFGRAEVAGTGD"
complement(7210..10254)
/gene="fscRIII"
complement(7210..10254)
/gene="fscRIII"
/notes="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="fscRIII"
/protein_id="AAQ82553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPPIELPRLDRDLAALS AVIGELSGRPAVTVTQEP
GLQGNLDLWAAAYATDAGRLVLSAHATPAEHEVRYGVVAQLLAENRALAPFLFTD
EQPGGLPGLNGLNAVDRPTLLVDRDQWLDPSLWLEALARLPAPVALLTSTT
GTALPVEHVSPTGLTATTIELAPLPTTSGGTAVLRAFAGAPGDPAFTEALABAT
RGVPAVHVLDLRFABAGVSPRADRIDPLRALTAEVVVDHATBALSDLRPDAVALR
ALAVCGDLLDFPLVITLAPHSVSESLRRLAASGLTILRDGHPRVQDAVVARVLE
EMPAADRAELYAFAGLAQRAADDQDILLILARPVGDPFMAVDTLIRGFTSALRGG
RDLVAYVYARLADDEFLADDRARIIEFQLASVEMVTAPTAAERLGLLIRATPRGGA
GLRARATDCLLGGDTRAAHLAGAIDSPAPAEPRRGPTGAPVPPPEAEPASTPT

```



```
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OSJNB0075N21"
/clone_lib="OSJNBb"

ORIGIN
Query Match 14.9%; Score 63.4; DB 8; Length 145539;
Best Local Similarity 55.8%; Pred. No. 16;
Matches 121; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 98 GCATCCAGAGACCGGCGCCGTCGGCGTGGAGGAGCGGAAGAACCGGCCGACG 157
Db 34685 GCAGCCAGAGTACTGTCTCCACAGGCTGGAGCGCTAGTCTAAGAGAGCGCCTCG 34744

QY 158 ACGCGGACCTCGTTCCCGCGCTCGCCCGAGGTCCTCGGCTCGACTCGACCTCGACGAAAGCCCTCG 217
Db 34745 ACGCGGCGGACGTCGTCCTCACTTCGAGCGCTCGGCGGCTCGGCGCCCGCCGCGCTCT 34804

QY 218 CGCGCGAGGTCGCGCCCGCGGTCACCCCGCGAGCGCCCAACCATGGAGCTGGAGC 277
Db 34805 TCATCGGCTGAGGCGCGCGGCGAGCGCGCGCCCGCGCCCGCTTCCGCTCA 34864

QY 278 AGGAATCGAGTGTCTCGCGACCGACCCCAAGCTGGA 314
Db 34865 CGCCGACGCGCTCGCTCAACGCGGCGAGCAGCA 34901

RESULT 13
SC0939115/c 276800 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 12/29.
DEFINITION AL939115 AL035640 AL035654 AL035707 AL049573 AL132674 AL345048
ACCESSION AL355832 AL359989 AL442120 AL596138 AL645882
VERSION AL939115.1 GI:24426505
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabbitts,R., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
TITLE Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21936410
PUBMED 12000953
REFERENCE 2 (bases 1 to 276800)
AUTHORS Bentley,S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA. E-mail: sdb@anger.ac.uk
COMMENT On or before Oct 30, 2002 this sequence version replaced
GI:4500374, GI:4490616, GI:4490978, GI:6138834, GI:7672242,
GI:7799503, GI:8894754, GI:10241774.
LOCATION/Qualifiers
FEATURES
source
1..276800
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
48..98
stem_loop

/note="Score 59: 21/22 ( 95%) matches, 0 gaps"
complement(146..652)
/gene="SCO3054"
/note="synonym: SCBAC19G2.09c"
complement(146..652)
/gene="SCO3054"
/note="SCBAC19G2.09c, conserved hypothetical protein, len:
168aa: similar to many in Streptomyces coelicolor eg.
TR:O9F2X0 (EMBL:AL392148) SCD20.09 hypothetical protein
(190 aa) fasta scores; opt: 250, Z-score: 285.6, 34.503%
identity (38.81% ungapped) in 171 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44519.1"
/db_xref="GI:14970940"
/db_xref="SPTRMBL:Q93J49"
/translation="MSRSPAPGGLVETLVNTLDLETGADSLDTEGRAPPLIED
GVTAARLRESLRTLLAHGHPFHRAVPLGLLAAAPLVAVDPAAGSAAAPAGE
GSUTARAAVAQAALVAGTWSRLKACEAAADCHWAYYDRSPAGRWCSNQVCGARAKM
RRYRAKES"
764..1123
/gene="SCO3055"
/note="synonym: SCBAC19G2.10"
764..1123
/gene="SCO3055"
/note="SCBAC19G2.10, conserved hypothetical protein, len:
119aa: similar to many in Streptomyces coelicolor e.g.
TR:O85701 (EMBL:AF072709) 3SCF60.11c conserved
hypothetical protein (131 aa) fasta scores; opt: 322,
Z-score: 387.9, 42.01% identity in 119 aa overlap and
downstream neighbouring CDS SCBAC19G2.11"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44520.1"
/db_xref="GI:14970941"
/db_xref="SPTRMBL:Q93J48"
/translation="MPVAELGVVLDCPPFPALARYADVLTGTVGEGDWDLRPLDG
GRALFAAPGFVPFPKAPAPDDSQPHLDLDVLDAAEKAVLELGARPLDAEDRTGT
FRYVADPAGHPFLCAC"
786..1114
/gene="SCO3055"
/note="Directly repeated at 11832..12181"
1153..1545
/gene="SCO3056"
/note="synonym: SCBAC19G2.11"
1153..1545
/gene="SCO3056"
/note="SCBAC19G2.11, conserved hypothetical protein, len:
aa: similar to many in Streptomyces coelicolor eg.
TR:Q9F3E9 (EMBL:AL450289) SC2H2.06 hypothetical protein
(150 aa) fasta scores; opt: 340, Z-score: 415.4, 46.032%
identity in 126 aa overlap and upstream neighbouring CDS
SCBAC19G2.10"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC44521.1"
/db_xref="GI:14970942"
/db_xref="SPTRMBL:Q93J47"
/translation="MAPAARFRSVVDCPPRELARFYAAGGTTDEADPDWVLQV
PGEPRLSFRAPDLTTPWPSPDRNACQPHLDFDGGATWAEMDAADRVILGARPLD
LEEDKDFWYADPAGHPFLCRIEHT"
1178..1527
/gene="SCO3056"
/note="Directly repeated at 11440..11768"
complement(1558..2688)
/gene="SCO3057"
/note="synonym: SCBAC19G2.12c"
complement(1558..2688)
/gene="SCO3057"
/note="SCBAC19G2.12c, possible dipeptidase, len: 376 aa:
repeated_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit

gene
CDS

repeat_unit
```

similar to many eukaryotic peptidases e.g. SW:P31430 (MDPL_RAT) rat renal dipeptidase (410 aa) fasta scores; opt: 440, Z-score: 474.0, 36.05% identity (38.906% unaligned) in 355 aa overlap. Also weakly similar to Brevibacillus borstelensis (307 aa) fasta scores; opt: 273, Z-score: 298.4, 28.159% identity (30.116% unaligned) in 277 aa overlap. Contains pfam match to entry PF01244 Renal dipeptidase. Renal dipeptidase. Also similar to neighbouring upstream CDS SCBAC19G2.13c
/codon_start=1
/transl_table=1
/product="putative dipeptidase"
/protein_id="CAC44522.1"
/db_xref="GI:14970943"
/db_xref="GOA:Q93J46"
/db_xref="STRMBL:Q93J46"
/translation="MADLQDDUHTGEAGGLDLPVPLAAEPPEPVAPVSDPDDEP LARAHVLAHPVADGNCGLPWALRHLPWYDELGESAVDTIDVPRLRGHVGLLWSL HLPESLDGPRVATLEQLDLVKTIVRAHPGEGLRAYDAGQDAIDRNCRIAVLPFPA GAAALDRIGLIRLHALGLRLVLTISGVSWASEAGLTFGEVVRNRLGLVWVADLSG AGAETVRRTPFAVSKAPALCTRGAAALRHPANLPDILLVELGAAGGLCMVPLTAEO T GPTVRVADHLHVRTVAGQSVGLSGTYDSCAAHPLGLDPSCYPRILVAELLRRWD EADVALLTGNVQRVLRRAAFTAKAAQLRREPSTATIADLG"
/genes="SCO3057"
/notes="Pfam match to entry PF01244 Renal dipeptidase, Renal dipeptidase, score 121.70, E-value 1.4e-32"
/complement(1562..2643)
/notes="Degenerately, directly repeated at 13490..14610"
complement(2808..4010)
/genes="SCO3058"
/notes="synonym: SCBAC19G2.13c"
complement(2808..4010)
/genes="SCO3058"
/note="SCBAC19G2.13c, possible dipeptidase, len: 400 aa: similar to many eukaryotic peptidases e.g. SW:P31429 (MDPL_RABIT) rabbit renal dipeptidase (410 aa) fasta scores; opt: 788, Z-score: 879.1, 42.432% identity (48.580% unaligned) in 403 aa overlap. Also weakly similar to TR:Q9KH70 (EMBL:AF268476) the most stable dipeptidase from Brevibacillus borstelensis (307 aa) fasta scores; opt: 450, Z-score: 506.1, 31.034% identity (36.486% unaligned) in 348 aa overlap. Contains pfam match to entry PF01244 Renal dipeptidase. Renal dipeptidase. Also similar to neighbouring downstream CDS SCBAC19G2.12c"
/codon_start=1
/transl_table=1
/product="putative dipeptidase"
/protein_id="CAC44523.1"
/db_xref="GI:14970944"
/db_xref="GOA:Q93J45"
/db_xref="STRMBL:Q93J45"
/translation="MTSLKARELLREFFVVDGHNLPALREQVRVYDLDDIADQ SHAHLDLARLSSGVGQVSVYVRSYDLPGAVATLEQIDCVRLRLDHPGELEAAL TAADMEARAGRIASLGAEGGSHIDSLATLALYALGVRYMTLTHNDNNVADSA TDPGVGGISLQGRVEMRNMREGLVDLSHVAATMRDALDTAPVIFSHSSSRV CDPRNI PDVLRLSRLSANGMAMTVFVKFVLQAVDVTAEADNNRAHGFHLDSSP EAMKHAPEERVPVATVSTVADHLHREVAGVHLGIGGYDGTFFPDGLGDV SGYPLNILLDRGWSQSLAKLTKWNAVRVLDAAEDVSRGLRAARGPSNATIEQLDG TAAEQPEG"
complement(2820..4010)
/genes="SCO3058"
/note="Pfam match to entry PF01244 Renal dipeptidase, Renal dipeptidase, score 280.40, E-value 2.3e-80"
complement(2836..3956)
/note="Degenerately, directly repeated at 12216..13297"
complement(4025..4567)
/genes="SCO3059"
/note="synonyms: pure, SCBAC19G2.14c"
complement(4025..4567)
/genes="SCO3059"
/EC_number="4.1.1.21"

/note="SCBAC19G2.14c, phosphoribosylaminimidazole carboxylase catalytic subunit PureE, len: 180aa: strongly similar to many eg. SW:Q45679 (PUR6_CORAM) phosphoribosylaminimidazole carboxylase catalytic subunit Pure from Corynebacterium ammoniagenes (177 aa) fasta scores; opt: 666, Z-score: 735.8, 65.541% identity (65.541% unaligned) in 148 aa overlap and TR:O80937 (EMBL:AC004684) putative phosphoribosylaminimidazole carboxylase from Ataxidopsis thaliana (645 aa) fasta scores; opt: 644, Z-score: 703.2, 60.870% identity (60.870% unaligned) in 161 aa overlap. Contains Pfam match to entry PF00731 AIRC, AIR carboxylase."
/codon_start=1
Query Match 14.7%; Score 62.6; DB 1; Length 276800;
Best Local Similarity 49.5%; Pred. No. 17;
Matches 161; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 24 GCGCGGCACATGGCGCCCATACGTCTCACCAGCGCGGAGCGCGGACTGACCAAGAG 83
Db 43436 GGGCGTCGAGAAGCGCGTGTGCGCGTCATCGAGCAGATCGCGCCGAGCTGTCGGCTA 43377
QY 84 CGAGTTGGCGAGCGGCATCCAGAAGGACCGGCGCCACCGTCGCGCGGTGGAGGACGGCAA 143
Db 43376 CGAGCGCACGAGCGGCGCTGATCGACGAGGCATGTTGACCTGAGCGCCACCGACAA 43317
QY 144 GAACCGCGCCGAGCGGAGCGGACCTCGTTGCCCGTGGCCAGGTGCTGGCGTGCACCT 203
Db 43316 CAAGGGCTCGTCGGCGCCAAACGCCATCTCGCGCTCTCCCTCGCGTGGCCACCGCCG 43257
QY 204 CGAGGAGCGCTCGCGCGCGCGGTCTGCGCGCGGCTCACCCCGCGAGCGGACCCCAAC 263
Db 43256 CTCGAGCGCGAGCGACTCGCGCTCTTCGCTTACCTGGGCGGCGCCGAAACGCGACCTGCT 43197
QY 264 CATGGACTCGAGGAGGAATCGAGCTGGTTCGCGACCGACCCCAAGCTGGAGGAGACAT 323
Db 43196 GCGCGTCCGATGATGAACATCTGAACGCGGCTCGACGCGGACTCCAACTGGACAT 43137
QY 324 GAGCGCGCATCATGCCCTAATC 348
Db 43136 CCAGGATTCATGATCGCCCGATC 43112
RESULT 14
SSBGL3GE SSBGL3GE 2081 bp DNA linear BCT 03-AUG-1994
LOCUS Streptomyces sp. QM-B814 (AtCC 11238) bg13 gene for beta-glucosidase.
DEFINITION beta-glucosidase.
ACCESSION Z29625
VERSION Z29625.1 GI:525288
KEYWORDS beta-glucosidase; bg13 gene.
SOURCE Streptomyces sp.
ORGANISM Streptomyces sp.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 2081)
AUTHORS Perez-Pons,J.A., Cayetano,A., Rebordosa,X., Lloberas,J., Guasch,A. and Querol,E.
TITLE A beta-glucosidase gene (bg13) from Streptomyces sp. strain QM-B814. Molecular cloning, nucleotide sequence, purification and characterization of the encoded enzyme, a new member of family 1 Glycosyl hydrolases
JOURNAL Eur. J. Biochem. 223 (2), 557-565 (1994)
MEDLINE 94333346
PUBMED 8055926
REFERENCE 2 (bases 1 to 2081)
AUTHORS Perez-Pons,J.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-1994) Josep A. Perez-Pons, Bioquímica i Biologia Molecular, Institut Biologia Fonamental & Dept. Biologia i Biol. Mol., Campus Bellaterra. U.A.B., Bellaterra, Barcelona. Catalunya, 08193 Barcelona, Spain
FEATURES Location/Qualifiers

```
source
1. 2081
/organism="Streptomyces sp."
/mol_type="genomic DNA"
/strain="QM-B814 (ATCC 11238)"
/db_xref="taxon:1931"
/clone="Clone pJC911"
repeat_region
134..173
/rpt_type=DIRECT
repeat_unit
134..148
repeat_unit
159..173
RBS
174..180
/standard_name="shine-dalgarno sequence"
/notes="putative"
186..1625
/gene="bgl3"
186..1625
/gene="bgl3"
186..1625
/gene="bgl3"
/EC_number="3.2.1.21"
/function="hydrolyzes cellobiose and other
beta-D-glucosides"
/standard_name="Streptomyces sp. bgl3 (sspbgl3)"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="beta-glucosidase"
/protein_id="CAA82733.1"
/db_xref="GI:581738"
/db_xref="GOA:O59976"
/db_xref="SPTREMBL:O59976"
/translation="NVPAAQQTATPDAALTFPEGLFNGSATASVOIEGAAREDEGRTP
SWDYTFPVRNGDTGDVAIDHWRDEVALMELGAYRFLSIAWRIQPTGR
GPAQKGDYFRLADELLAKQIPVATLYHWDLPFELNPGWPFRTAFRABIAA
IAADALGRVKTWTTLNFWPCSGIFLGGVGHVGPRTDPAALRAAHNLHGLHGLAVQ
ALRDLPLDAQCSVTNLHVRPLTDSADADAVRRIIDALANRVFTGMLQAYPEDL
VKDTAGLTDSFVRGDLRLAHOKLDFLGVNYSPTLVSEADSGTHSDGHSRSHS
PMPGADRVAFHPPGPE""AMGWAVDPGSLVELLRELSDFPALPLVITENGAAPHDYA
DEGNVDNPERIAYVRDHLAAVHRAIKDGSVGRGFLNLSLLDNFENWAGYSKRFQAVY
VDYPTGTRIPKASRWYAEVARTGLVPTA"
1676..1719
/terminator
/notes="Putative"

ORIGIN
Query Match 14.5%; Score 61.6; DB 1; Length 2081;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 154; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 37 GCGGCATAGTCTCTACCCGCGGAGCGCGGCGGACTGACCAAGAGCGAGTGGCCAGG 96
DB 651 GCGGAGTACGCGGCATCGCGCGGAGCGCCCTCGCGGACCGGTGAAGACCTGGACCACC 710
QY 97 CGCATCCAGAGACCGGGCCACCGTCGCGCGGTGGAGGACGGGAGAACCGGCCGAC 156
DB 711 CTCAAGAGCCCTGTGTGACGCGGTTCTCTCGGTACGCGTCCGCGGTGACACGCCCGCGC 770
QY 157 GAGCGGACCTGTTCCCGCGTGCAGAGTCTCGGCTCGACCTCGACGAAGCCCTC 216
DB 771 CGACCGACCGGTGCGCCCTCGCGCGCGCACCACTCACTCGGCGCACGCGCTG 830
QY 217 GCGCGCGAGGTCTGCGCGCGCGCGGTCAACCCGCGGAGCGACCCCAACCAATGACCTGGAC 276
DB 831 GCGGTCCAGCGCTGCGGACCGCCCTCCCGCGGACGCGCGCTCGTCTCGTCACTCAAC 890
QY 277 GAGGAATCAGCTGGTTCGCGACCGCCCAAGCTGGAGGACATGAAGCGGCGCATC 336
DB 891 ATCCACACGTACGGCGGTCTACCGAGCGGCGGACGCGCGGTCCGCGGATC 950
QY 337 ATCGCCCT 344
DB 951 GACGGGCT 958
```

RESULT 15
BD179992

```
LOCUS BD179992 1233 bp DNA linear PAT 15-MAY-2003
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION BD179992
VERSION BD179992.1 GI:30790910
KEYWORDS JP 2002325574-A/483.
SOURCE Thermus thermophilus
ORGANISM Thermus thermophilus
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE
1 (bases 1 to 1233)
Kuramitsu, N. and Yokoyama, S.
Highly thermophilic bacterium-derived protein and gene encoding it
Patent: JP 2002325574-A 483 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
OS Thermus thermophilus
PN JP 2002325574-A/483
PD 12-NOV-2002
PF 23-FEB-2001 JP 2001116171
PI NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA
PC C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12N9/88,C12P21/02//(C12N9/88,C12R1:01),(C12N15/09,C12R1:01),
PC (C12P21/02,C12R1:01),C12N15/00,C12N15/00,C12N5/00,
PC C12R1:01)
CC Highly thermophilic bacterium-derived protein and gene CC
FH Key Location/Qualifiers
FT CDS (1)..(1233).
FEATURES
source
1..1233
/organism="Thermus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:274"
ORIGIN
Query Match 14.4%; Score 61.4; DB 6; Length 1233;
Best Local Similarity 47.5%; Pred. No. 1.1e+02;
Matches 182; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 2 TGCACAACACACCGGGCTGGGGCGGCGACATGGCGCGCATAGTCTCACCGCCGCG 61
DB 257 TGCACAACACCGGTGCGGGAGTGGGGTACACCCCGGCCAAGTACGGTTTTCACGCCGACA 316
QY 62 AGCGCGCGGACTGACCAAGAGCGAGTTGGCCAGCGCATCCAGAAGACCGGGCCACCG 121
DB 317 CTGCGCGCTCTCACCGCCATAGACGAGCAGTCCCGCGACATCGCCGGGGGTCAACC 376
QY 122 TCGCCCGGTGGAGGACGGCAAGAACCGGCGCGGACGACGCGGACCTGTTGCCCGGTG 181
DB 377 TCTCTACGAGTGGCGGTCTCTCAAGTCCACCGACCCCTGACCGGGTGGGGCCGCGG 436
QY 182 CCGAGGTCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCAGGTCTGCGCCCGCGCG 241
DB 437 ACCAGGGGCTCATGTTTGGTTACGCCACCGACGAGACCCCGAGCTATGCCCTTCCCA 496
QY 242 TCACCCCGCCAGACACCCCAACCATGGACCTGGACGAGGAATCGAGTGTCCGACCG 301
DB 497 TCACCTTGGCCACCGCTCACCATGCGCTCGCGAGGTGCGGAAGACGGGGTCTTGC 556
QY 302 ACCCAAGCTGGAGGACATGAGCGGCGCATATCGCCCTAATCCTGAGCGCGGTG 361
DB 557 CTAACCTCGGCGCGGACGCGCAAGCCAGGTCAACCGTGTCTACAGGGGACAGCCCC 616
QY 362 AGCGCGCAAGCGCGCGCGCATC 384
DB 617 TTTACGTCAGACCGTGTGTGTC 639
```

Search completed: August 6, 2004, 11:53:10
Job time : 1961.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 253.157 Seconds
(without alignments)
7148.662 Million cell updates/sec

Title: US-09-855-340A-2
Perfect score: 426
Sequence: 1 atgcgaacacacccgggct.....acctttccggcgagctga 426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	426	100.0	426	6	AAD25932
2	426	100.0	4388	5	AAS08694
3	66	15.5	1575	2	AAQ12342
4	64.4	15.1	1056	3	AAC55832
5	64.4	15.1	1056	9	ADE10251
6	64.4	15.1	53500	3	AAC55842
7	64.4	15.1	53500	9	ADE10261
8	60.6	14.2	58857	3	AAAS58471
9	56.8	13.3	972	7	ABZ66792
10	56.8	13.3	45055	7	ABZ66808
11	56.2	13.2	2640	7	ADA69934
12	56.2	13.2	2640	9	ADC08080
13	56.2	13.2	113193	7	AAD54645
14	56	13.1	662	3	AAF12486
15	55.4	13.0	88421	6	AAL40781
16	55	12.9	4770	7	AAL61172
17	55	12.9	82746	7	AAL61224
18	54.2	12.7	1299	7	ABZ71158
19	54.2	12.7	36321	7	ABZ71131
20	53.8	12.6	861	7	ACA03525
21	53.8	12.6	861	7	ACC78498
22	53	12.4	699	7	ADA70128
23	53	12.4	3072	7	ACA37796

24	53	12.4	5811	6	ABS78655
25	52.8	12.4	1266	7	ACA42530
26	52.8	12.4	8169	7	AAV28609
27	52.8	12.4	109519	5	AAS08693
28	52.4	12.3	125401	4	AAD17186
29	52	12.2	29870	7	AAD36874
30	51.6	12.1	1185	7	ADA69780
31	51.6	12.1	17083	7	ABZ58812
32	51.4	12.1	2000	7	ADA71938
33	51.4	12.1	27541	4	AAD17185
34	51.2	12.0	861	9	ACA03527
35	51.2	12.0	861	9	ADC13260
36	51.2	12.0	1389	7	ADA71184
37	51	12.0	1527	7	ACA37879
38	51	12.0	44377	2	AAT80414
39	51	12.0	44377	2	AAT78508
40	50.8	11.9	543	7	ACA37764
41	50.8	11.9	65140	4	AAD17184
42	50.6	11.9	1047	7	ABZ66673
43	50.6	11.9	31263	7	ACA37577
44	50.4	11.8	1224	3	AAC55788
45	50.4	11.8	1224	9	ADE10207

ALIGNMENTS

RESULT 1
AAD25932
ID AAD25932 standard; DNA; 426 BP.
XX
AC AAD25932;
XX
DT 26-MAR-2002 (first entry)
XX
DE Micromonospora carbonacea pMLP1 excisionase DNA.
XX
KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
KW site-specific integration; hybrid antibiotic; metabolic product;
KW secondary metabolic pathway; ds.
XX
OS Micromonospora carbonacea.
XX
PN WO200187936-A2.
XX
PD 22-NOV-2001.
XX
PF 15-MAY-2001; 2001WO-US015760.
XX
PR 17-MAY-2000; 2000US-0204670P.
XX
SC (SCHE) SCHERING CORP.
XX
Hosted TJ, Horan AC;
XX
WPI; 2002-082983/11.
XX
Novel polynucleotides encoding integrase, excisionase and an integrase
attachment site isolated from a lysogenic phage pMLP1, useful for
transforming an actinomycete.
XX
Claim 5; Page 32-33; 34pp; English.
XX
The present invention relates to novel polynucleotides encoding integrase
(int) and excisionase (xis) and an integrase attachment site (attP) which
are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
Micromonospora carbonacea var. africana. Polynucleotides of the invention
are useful for transforming an actinomycete with a vector. They are also
useful for creating vectors for site-specific integration into host
chromosomes. The integrating vectors are used to express actinomycete
genes, manipulate secondary metabolic pathways and create new metabolic
products such as hybrid antibiotics. The present sequence is pMLP1
excisionase DNA from Micromonospora carbonacea

```
XX SQ Sequence 426 BP; 82 A; 157 C; 143 G; 44 T; 0 U; 0 Other;
Query Match 100.0%; Score 426; DB 6; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.3e-64;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGAACACACCGGGGCTGGGGCGGCGGCACATGGCGCGCATACGTCCTCACCGCCGC 60
Db 1 ATGGCGAACACACCGGGGCTGGGGCGGCGGCACATGGCGCGCATACGTCCTCACCGCCGC 60

Qy 61 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGCGCGCATCCAGAAAGACCGGGCCACC 120
Db 61 GAGCGCCCGGACTGACCAAGAGCGAGTTGGCCAGCGCGCATCCAGAAAGACCGGGCCACC 120

Qy 121 GTGCGCGGTTGGAGGACGACAGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGC 180
Db 121 GTGCGCGGTTGGAGGACGACAGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGC 180

Qy 181 GCCCAGGTGCTGGGCTTCGACCTCGACGAAGACCCCTCGCGCGCGCGCGCGCGCGC 240
Db 181 GCCCAGGTGCTGGGCTTCGACCTCGACGAAGACCCCTCGCGCGCGCGCGCGCGC 240

Qy 241 GTACCCCGCGGAGGACCCCAACCATGGACTGGAGGAGGAATCGAGTGTCCGCACC 300
Db 241 GTACCCCGCGGAGGACCCCAACCATGGACTGGAGGAGGAATCGAGTGTCCGCACC 300

Qy 301 GACCCCAAGCTGGACGAGGACATGAAGCGCGCGCATATCGCCCTAATCTGGAGCGCGGT 360
Db 301 GACCCCAAGCTGGACGAGGACATGAAGCGCGCGCATATCGCCCTAATCTGGAGCGCGGT 360

Qy 361 GAGCGCGACAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 361 GAGCGCGACAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420

Qy 421 AGCTGA 426
Db 421 AGCTGA 426

RESULT 2
ID AAS08694
XX AAS08694 standard; DNA; 4388 BP.
AC AAS08694;
XX
DT 11-SEP-2003 (revised)
DT 26-SEP-2001 (first entry)
XX
XX Micronospora DNA encoding integrase enzymes.
XX Evernimycin; antibiotic; bottle-neck gene; orthomycin; fermentation;
KW integrase; ds.
XX
XX Micronospora sp. ATCC 39149.
XX
XX Key Location/Qualifiers
XX CDS 963..1387
XX /*tag= a
XX /product= "Integrase #1"
XX 1394..2572
XX /*tag= b
XX /product= "Integrase #2"
XX 2570..2799
XX /*tag= c
XX /note= "AttB/AttP region of integrase action"
XX 27114..2715
XX /*tag= d
XX /label= Insertion_juncture
XX /note= "Site of integrase activity"
XX
XX W0200151639-A2.
```

```
PD 19-JUL-2001.
XX
XX 12-JAN-2001; 2001WO-US001187.
XX
XX 12-JAN-2000; 2000US-0175751P.
XX
XX (SCHE ) SCHERING CORP.
XX
XX Hosted TJ, Horan AC, Wang TX;
XX
XX WPI; 2001-442147/47.
XX
XX P-PSDB; AAU04900, AAU04912.
XX
XX New nucleic acid molecules encoding evernimycin pathway gene products,
XX useful for improving yields of evernimycin, to produce new
XX evernimycin and as probes to identify homologous sequences.
XX
XX Claim 26; Fig 7; 109pp; English.
XX
XX The sequence encodes 2 integrases which permit site specific integration
XX of a vector into an actinomycete, especially a Micromonospora, genome.
XX The invention relates to nucleic acids and vectors comprising a M.
XX carbonacea evernimycin biosynthetic pathway resistance gene product
XX useful for selection of a transfected or transformed host cell. An
XX integrative version of the vector is useful for introducing a
XX evernimycin pathway gene (a bottle-neck gene) into an actinomycete of
XX the genus Micromonospora. The DNA encoding the biosynthetic proteins is
XX useful for synthesizing novel evernimycin-related compounds, arising
XX from modifications of the DNA sequence designed to change glycosyl and
XX modified orsellinic acid groups contained in evernimycin, for
XX expressing functional or mutant evernimycin biosynthetic enzyme for
XX evaluation, diagnosis and preferably biosynthesis of evernimycin or
XX other secondary metabolic products, improving the yield of evernimycins
XX and to produce novel evernimycins and also as a hybridisation probe to
XX identify homologous sequences. The encoded polypeptides are useful for
XX combinatorial biosynthesis to generate libraries of orthomycins, e.g.
XX evernimycin analogues/homologues and drug discovery. The DNA encoding
XX the integrase allows for increasing a given gene dosage. The integrative
XX vector can be used to permanently integrate copies of a heterologous gene
XX of choice into chromosomes of different hosts and to integrate genes
XX which increase the yield of known products or to generate novel products
XX such as hybrid antibiotics or other novel secondary metabolites. The
XX vector can also be used to integrate antibiotic resistance genes in order
XX to carry out bioconversions with compounds to which the strain is
XX normally sensitive and is thus useful in fermentation processes involving
XX e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS
XX field)
XX
XX Sequence 4388 BP; 714 A; 1468 C; 1506 G; 694 T; 0 U; 6 Other;
Query Match 100.0%; Score 426; DB 5; Length 4388;
Best Local Similarity 100.0%; Pred. No. 1e-64;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGAACACACCGGGGCTGGGGCGGCGGCACATGGCGCGCATACGTCCTCACCGCCGC 60
Db 963 ATGGCGAACACACCGGGGCTGGGGCGGCGGCACATGGCGCGCATACGTCCTCACCGCCGC 1022

Qy 61 GAGCGCGCGGACTGACCAAGAGCGAGTTGGCCAGCGCGCATCCAGAAAGACCGGGCCACC 120
Db 1023 GAGCGCGCGGACTGACCAAGAGCGAGTTGGCCAGCGCGCATCCAGAAAGACCGGGCCACC 1082

Qy 121 GTCGCGCGGTTGGAGGACGCGCAAGAACCGGCGCGCGCGCGCGCGCGCGCGCGC 180
Db 1083 GTCGCGCGGTTGGAGGACGCGCAAGAACCGGCGCGCGCGCGCGCGCGCGCGCGC 1142

Qy 181 GCCCAGGTGCTGGGCTTCGACCTCGACGAAGACCCCTCGCGCGCGCGCGCGCGC 240
Db 1143 GCCCAGGTGCTGGGCTTCGACCTCGACGAAGACCCCTCGCGCGCGCGCGCGCGC 1202

Qy 241 GTCACCCCGCGGAGGACGCGCAAGAACCGGCGCGCGCGCGCGCGCGCGCGCGC 300
Db 1203 GTCACCCCGCGGAGGACGCGCAAGAACCGGCGCGCGCGCGCGCGCGCGCGCGC 1262
```

QY 301 GACCCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAAATCTCTGGAGCGCCCT 360
 |||||
 Db 1263 GACCCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAAATCTCTGGAGCGCCCT 1322
 |||||
 QY 361 GACCGGACAAAGCGCGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCCGG 420
 |||||
 Db 1323 GACCGGACAAAGCGCGCGGCGATCGAGGAAACCAAGCGGCTCATCGACCTGTTCCGCCGG 1382
 |||||
 QY 421 AGCTGA 426
 |||||
 Db 1383 AGCTGA 1388
 |||||

RESULT 3
 AAQ12342
 ID AAQ12342 standard; DNA; 1575 BP.
 XX AC
 AC AAQ12342;
 XX DT 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 12-SEP-1991 (first entry)
 XX DE Actinomycete Phospholipase D gene.
 XX DE
 KW lecithin; actinomycetes; phospholipid; ss.
 XX OS Streptomyces chromofuscus; A 0848.
 XX FH Key Location/Qualifiers
 FT sig_peptide 1..42 /*tag= a
 FT mat_peptide 43..1572 /*tag= b
 FT /product= "phospholipase D"

XX EP435725-A.
 PN
 XX
 XX 03-JUL-1991.
 PD
 XX 11-DEC-1990; 90EP-00403529.
 PF
 XX 15-DEC-1989; 89JP-00325355.
 PR (TOXN) TOYO JOZO KK.
 PA (ASAH) ASAH KASEI KOGYO KK.
 XX Yoshioka I, Mizoguchi J, Takahara M, Imamura S, Bappu T;
 PI Horinouchi S;
 PI WPI; 1991-195237/27.
 XX P-PSDB; AAR12786.
 DR
 XX Cloned DNA encoding phospholipase D - having enzymatic activity and
 PT produced efficiently in pure form.
 PT Claim 2; Fig 3; 32pp; English.
 XX PS
 XX This sequence was isolated from a gene library prepared from the
 CC actinomycete Streptomyces chromofuscus A 0848 strain. It was present in
 CC plasmid pcPD1. See also AAQ12341. (Updated on 25-MAR-2003 to correct PA
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)
 XX Sequences 1575 BP; 228 A; 627 C; 514 G; 206 T; 0 U; 0 Other;
 SQ
 Query Match 15.5%; Score 66; DB 2; Length 1575;
 Best Local Similarity 48.4%; Pred. No. 0.0069;
 Matches 183; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
 QY 7 AACACACCGGGTGGGCGCGGCACATGGCGCGATAGTCTACCGCGCGCGAGCGC 66
 |||||
 Db 646 AAGGCGCGCGTCTCGCGCATCTGGGACGACACCGAGTTCGCCGACAAACGCGCTGGTCCGCG 705
 |||||

QY 67 GCCGACTGACCAAGAGGAGTTCGCCAGGCGCATCCAGAAAGACCGGCGCACCGTCGCC 126
 |||||
 Db 706 GCGCGGTGAACCAACACACGAGGGCGCGAGGGCACCTGGTCGCGCCCGTCAGGCGCGCGCC 765
 |||||
 QY 127 CGGTGGAGGACGCGCAAGAACCGGCCGACGACGCGGACCTTCGTTGCCCGGTGCGCCAG 186
 |||||
 Db 766 AAGCAGGCGCTACTTCGAGTGGATGCGCGTGGCGCCCGCCATCGCGGGACACCACTACGG 825
 |||||
 QY 187 GTGCTCGGCTCGACCTCGACGAAGCGCTCGCCGCGCGAGGTCTGCGGCCCGGGGTACCC 246
 |||||
 Db 826 CGGCTGCGCTTCGGCAAGCTCGCGACCTCTCCCTGTGGACCTGCGCTCTCCGCTCC 885
 |||||
 QY 247 CGCGCAGCGACCCCAACCATGACCTGACGAGGAATCGAGCTGGTCCGACCGACCCC 306
 |||||
 Db 886 CAGCAGGCGCTCCAGCGCCAGCGGTTCGGTGAACGACCGGACCGTACGCTCACCGGCGCG 945
 |||||
 QY 307 AAGCTGGACGAGGACATGAAGCGCGCATCATCGCCCTAAATCTCTGGAGCGCGGTAGCGC 366
 |||||
 Db 946 GCGCAGCTCGACTGGCTCAGGCGGCGCTGAAGGCTCGGACACACAGTGGCGGTGGTC 1005
 |||||
 QY 367 GACAGGCGCGCGGATC 384
 |||||
 Db 1006 GGCAACTCCGTGATGATC 1023
 |||||
 RESULT 4
 AAC55832
 ID AAC55832 standard; DNA; 1056 BP.
 XX AC
 AC AAC55832;
 XX DT 19-JAN-2001 (first entry)
 XX DE S. lavendulae MmcR encoding DNA sequence.
 XX KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide; ds.
 XX OS Streptomyces lavendulae.
 XX PN WO200053737-A2.
 XX 14-SEP-2000.
 XX 10-MAR-2000; 2000WO-US006394.
 XX 12-MAR-1999; 99US-00266965.
 XX (MINU) UNIV MINNESOTA.
 PA (SHER/) SHERMAN D H.
 PA (MAOY/) MAO Y.
 PA (VARO/) VAROGLU M.
 PA (HEM/) HE M.
 PA (SHEL/) SHELTON P C.
 XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 PI WPI; 2000-601980/57.
 DR P-PSDB; AAB32531.
 XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
 PT the molecular basis of mitosome ring system biosynthesis.
 PS Disclosure; Page 281; 399pp; English.
 XX This invention relates to isolated and purified nucleic acid molecules
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
 CC natural products that contain a variety of functional groups, including
 CC amino benzoquinone and axiridine ring systems. The S. lavendulae

CC Mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning
 CC 55kb of DNA. The invention includes an expression cassette comprising a
 CC mitomycin biosynthetic gene operably linked to a promoter, and host cells
 CC transformed with the cassette. The nucleotide, and protein sequences and
 CC the transformed host cells of the invention result in antiasthmatic,
 CC anti-inflammatory, cytostatic, immunomodulatory, and antibiotic
 CC activities. The nucleotide sequences are used to elucidate the molecular
 CC basis for the biosynthesis of the mitomycin gene system, as well as to
 CC engineer the biosynthesis of novel natural products, e.g. antibiotics,
 CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other disease involving respiratory inflammation, or
 CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides
 CC or insecticides) as well as biopolymers, e.g., in packaging or biomedical
 CC applications, or to engineer PHA monomer synthases. Sequences AAC55782-
 CC C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin
 CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences
 CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR
 CC primers used in the cloning of the mitomycin biosynthetic genes
 XX
 SQ Sequence 1056 BP; 152 A; 427 C; 338 G; 139 T; 0 U; 0 Other;

Query Match 15.1%; Score 64.4; DB 3; Length 1056;
 Best Local Similarity 51.0%; Pred. No. 0.014;
 Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAGAGACCGGCGCCACCGTCGGCGGTGGGAGGACGGCAAGAACCGGC 151
 Db 683 CCGGCGCGCGCTTCGGGACCGGTGGAGATCTCTCGCGGCGACTTCTTCGAGACCATCC 742

QY 152 CCGAGGACGGGACCTGTTGCCCGGTGCGCCAGGTGTCGCCCTCGACCTCGAGGAG 211
 Db 743 CCGAGGCGCGCGACTTACTCATCAAGCACGTCTGACGACTGGGACGACGACG 802

QY 212 CCTCTCGCGCGCGAGGTCTGCGCGCGGTACCCCGCGGCGCATGAAGCGGACTCCGCGCTCTGTGTA 862
 Db 803 TCGTACGATCTCTCGCGCGGTACCCCGCGGCGCATGAAGCGGACTCCGCGCTCTGTGTA 862

QY 272 TGGACGAGGAATCGAGCTGTCGACCGACCCAGCTGGACGAGGACATGAAGCGGC 331
 Db 863 TCGACAACTCATCGAGGACGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGCTGC 922

QY 332 GCATCATCGCCCTAATCTCTGAGCGCGTGTGAGCGCGCAAGCGCGCGCGATCGAGGA 389
 Db 923 TCGTCTCTGTCGGCGCGCGCAAGCGCTCGGAGAGCGAATTCGCGCGCTGCTGGAGAA 980

RESULT 5
 ADE10251
 ID ADE10251 standard; DNA; 1056 BP.

XX AC ADE10251;
 XX
 XX 29-JAN-2004 (first entry)

XX S. lavendulae mitomycin biosynthetic protein MmcR gene.

XX Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;
 XX tumour hypoxia; cytostatic; anti-tumour agent; cancer; gene.

XX Streptomyces lavendulae.

XX US2003134398-A1.

XX 17-JUL-2003.

XX 12-SEP-2001; 2001US-00953348.

XX 12-SEP-2001; 2001US-00953348.

XX (SHER/) SHERMAN D H.

XX (NAOY/) MAO Y.

XX (VARO/) VAROGLU M.

PA (HEMV/) HE M.
 PA (SHEL/) SHELTON P.

XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;

XX WPI; 2003-863498/80.

XX P-PSDB; ADE10303.

XX New nucleic acid molecule comprising a sequence having mitomycin
 PT biosynthetic gene cluster, useful for enhancing production of
 PT antibiotics.

XX Disclosure; SEQ ID NO 66; 308pp; English.

XX The invention relates to an isolated and purified nucleic acid molecule
 CC comprising a sequence having mitomycin biosynthetic gene cluster, or its
 CC variant or fragment. Also included are an expression cassette comprising
 CC the novel nucleic acid molecule (operably linked to a promoter functional
 CC in a host cell), a recombinant bacterial host cell in which at least a
 CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene
 CC cluster is disrupted (resulting in a recombinant host cell that produces
 CC altered levels of mitomycin relative to a corresponding nonrecombinant
 CC bacterial host cell), introducing exogenous DNA into a refractory
 CC Streptomyces strain, identifying a nucleic acid molecule that is related
 CC to at least a portion of a nucleic acid molecule comprising a mitomycin
 CC gene cluster, preparing a compound or its salt from the recombinant host
 CC cell and a product produced by the recombinant host cell. The nucleic
 CC acid encodes a MitR, MitS, MitQ, MitP, MitO, MitN, MitM, MitL, MitK,
 CC MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or
 CC MmcA-MmcY. The nucleic acid is useful for enhancing production of
 CC mitomycin antibiotics, which induce apoptosis and hence are useful as
 CC anti-tumour (via tumour hypoxia) agents and are useful in treating
 CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The
 CC present sequence is a gene encoding a mitomycin biosynthetic protein of
 CC the invention.

XX Sequence 1056 BP; 152 A; 427 C; 338 G; 139 T; 0 U; 0 Other;

Query Match 15.1%; Score 64.4; DB 9; Length 1056;
 Best Local Similarity 51.0%; Pred. No. 0.014;
 Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 92 CCAGGCGCATCCAGAGAGACCGGCGCCACCGTCGGCGGTGGGAGGACGGCAAGAACCGGC 151
 Db 683 CCGGCGCGCGCTTCGGGACCGGTGGAGATCTCTCGCGGCGACTTCTTCGAGACCATCC 742

QY 152 CCGAGGACGGGACCTGTTGCCCGGTGCGCCAGGTGTCGCCCTCGACCTCGAGGAG 211
 Db 743 CCGAGGCGCGCGACTTACTCATCAAGCACGTGTGACGACTGGGACGACGACG 802

QY 212 CCTCTCGCGCGCGAGGTCTGCGCGCGGTACCCCGCGGCGCATGAAGCGGACTCCGCGCTCTGTGTA 862
 Db 803 TCGTACGATCTCTCGCGCGGTACCCCGCGGCGCATGAAGCGGACTCCGCGCTCTGTGTA 862

QY 272 TGGACGAGGAATCGAGCTGTCGACCGACCCAGCTGGACGAGGACATGAAGCGGC 331
 Db 863 TCGACAACTCATCGAGGACGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGCTGC 922

QY 332 GCATCATCGCCCTAATCTCTGAGCGCGTGTGAGCGCGCAAGCGCGCGCGATCGAGGA 389
 Db 923 TCGTCTCTGTCGGCGCGCGCAAGCGCTCGGAGAGCGAATTCGCGCGCTGCTGGAGAA 980

RESULT 6
 AAC55842
 ID AAC55842 standard; DNA; 53500 BP.

XX AC AAC55842;

XX 19-JAN-2001 (first entry)

XX Complete nucleotide sequence of the mitomycin biosynthetic genes.

KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW fungicide; pesticide; ds.
XX
OS Streptomyces lavendulae.
XX WO200053737-A2.
XX PD 14-SEP-2000.
XX PF 10-MAR-2000; 2000WO-US006394.
XX PR 12-MAR-1999; 99US-00266965.
XX (MINU) UNIV MINNESOTA.
PA (SHER/) SHERMAN D H.
PA (MAO/) MAO Y.
PA (VARO/) VAROGLU M.
PA (HEMM/) HE M.
PA (SHEL/) SHELTON P C.
XX
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX WPI; 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT the molecular basis of mitosome ring system biosynthesis.
XX
XX Disclosure; Fig 26; 39pp; English.
XX
CC This invention relates to isolated and purified nucleic acid molecules
CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC natural products that contain a variety of functional groups, including
CC amino benzquinone and axiridine ring systems. The S. lavendulae
CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes spanning a
CC 55kb of DNA. The invention includes an expression cassette comprising a
CC mitomycin biosynthetic gene operably linked to a promoter, and host cells
CC transformed with the cassette. The nucleotide, and protein sequences and
CC the transformed host cells of the invention result in antiasthmatic,
CC antiinflammatory, cytostatic, immunomodulatory, and antibiotic
CC activities. The nucleotide sequences are used to elucidate the molecular
CC basis for the biosynthesis of the mitosome ring system, as well as to
CC engineer the biosynthesis of novel natural products, e.g. antibiotics,
CC anti-inflammatory agents, anti-cancer agents, immune-enhancers,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other disease involving respiratory inflammation, or
CC cholesterol-lowering agents or as crop protection agents (e.g. fungicides
CC or insecticides) as well as biopolymers, e.g., in packaging or biomedical
CC applications, or to engineer PHA monomer syntheses. Sequences AAC55782-
CC C55881, AAC55815-C55849 and AAB32485-B32542 represent mitomycin
CC biosynthetic gene cluster DNA sequences and encoded proteins. Sequences
CC AAC55812-C55814, AAC55850-C55856 and AAC55862-C55869 represent PCR
CC primers used in the cloning of the mitomycin biosynthetic genes
XX
XX Sequence 53500 BP; 7481 A; 19740 C; 19126 G; 7153 T; 0 U; 0 Other;
Query Match 15.1%; Score 64.4; DB 3; Length 53500;
Best Local Similarity 51.0%; Pred. No. 0.0091;
Matches 152; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
XX
QY 92 CCAGGGGCTCCAGAGGACCGGGCCACCGTGGCGGGTGGAGGACGCGCAAGACCGGC 151
DB 43516 CCGCGCCGGCTCCGCGGACCGGTGGAGATCTGCCCGGCACTTCTTCGAGACCATCC 43575
QY 152 CCAGACACGGGACCTCGTGGCCCGGTGCCCGAGGTGTGGCGCTTCGACCTCGAGGAG 211
DB 43576 CCGAGCGGCGGACGCTACCTCATCAAGCAGCTGCTGCACGACTGGGACGACGACG 43635
QY 212 CCTTCGCGCGGAGCTGTGGCGCCCGGTACCCCGGACGACCGCCCAACATGGACC 271
DB 43636 TGTACGCATCTCCCGCGGATCGCCACCGCATGAAGCCGCACTCCCGGCTCTCTGGTCA 43695

QY 272 TGGACGAGGAATCGAGCTGGTCCGACCGACCCAGCTGGAGGAGCATGAAGCGGC 331
DB 43696 TCGACCACTCATCGACGAGCGCGCGCATGAGCTCTTCTGCTGACTGTGCTGC 43755
QY 332 GCATCATCGCCCTAATCTCGAGCGCGGTGAGCGCGCAAGCGCGCGCATCGAGGA 389
DB 43756 TCGTCTCTGTCGGCGGCGCGAACGCTCGGAGAGCGAATTCGCGCGCTGCTGGAGAA 43813
RESULT 7
ADE10261
ID ADE10261 standard; DNA; 53500 BP.
XX
AC ADE10261;
XX
DT 29-JAN-2004 (first entry)
XX
DE S. lavendulae mitomycin biosynthetic genes complete sequence.
XX
KW Mitomycin biosynthetic protein; mitomycin C; antibiotic; MC; apoptosis;
KW tumour hypoxia; cytostatic; anti-tumour agent; cancer; ds.
XX
OS Streptomyces lavendulae.
XX
PN US2003134398-A1.
XX
PD 17-JUL-2003.
XX
PF 12-SEP-2001; 2001US-00953348.
XX
PR 12-SEP-2001; 2001US-00953348.
XX
PA (SHER/) SHERMAN D H.
PA (MAO/) MAO Y.
PA (VARO/) VAROGLU M.
PA (HEMM/) HE M.
PA (SHEL/) SHELTON P.
XX
PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon P;
XX WPI; 2003-863498/80.
XX
XX New nucleic acid molecule comprising a sequence having mitomycin
PT biosynthetic gene cluster, useful for enhancing production of
PT antibiotics.
XX
PS Disclosure; SEQ ID NO 76; 308pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid molecule
CC comprising a sequence having mitomycin biosynthetic gene cluster, or its
CC variant or fragment. Also included are an expression cassette comprising
CC the novel nucleic acid molecule operably linked to a promoter functional
CC in a host cell), a recombinant bacterial host cell in which at least a
CC portion of a nucleic acid molecule comprising mitomycin biosynthetic gene
CC cluster is disrupted (resulting in a recombinant host cell that produces
CC altered levels of mitomycin relative to a corresponding nonrecombinant
CC bacterial host cell), introducing exogenous DNA into a refractory
CC Streptomyces strain, identifying a nucleic acid molecule that is related
CC to at least a portion of a nucleic acid molecule comprising a mitomycin
CC gene cluster, preparing a compound or its salt from the recombinant host
CC cell and a product produced by the recombinant host cell. The nucleic
CC acid encodes a MitT, MitS, MitR, MitQ, MitP, MitD, MitC, MitB, MitA and/or
CC MitK, MitJ, MitI, MitH, MitG, MitF, MitE, MitD, MitC, MitB, MitA and/or
CC MmcA-MmcY. The nucleic acid is useful for enhancing production of
CC mitomycin antibiotics, which induce apoptosis and hence are useful as
CC anti-tumour (via tumour hypoxia) agents and are useful in treating
CC cancer. The gene cluster was isolated from Streptomyces lavendulae. The
CC present sequence is an operon or gene cluster encoding the mitomycin
CC biosynthetic proteins of the invention.
XX
SQ Sequence 53500 BP; 7481 A; 19740 C; 19126 G; 7153 T; 0 U; 0 Other;

Query Match	15.1%	Score 64.4	DB 9	Length 53500
Best Local Similarity	51.0%	Pred. No. 0.0091		
Matches 152	Conservative 0	Mismatches 146	Indels 0	Gaps 0
QY	92	CCAGGCGCATCCAGAAAGACCGGGCCACCGTCTCGCGCGGTGGGAGGACGGCAAGAACCGGC	151	
Db	43516	CCGCGCGCGCGCTTCGCGGACCGGTGGAGATCTCGCGGCGACTTCTTCGAGACCATCC	43575	
QY	152	CCGACGACGCGGACCTTCGTTGCCCGCGTCCGCCAGTGTCTCGGCTTCGACCTCGACGAAG	211	
Db	43576	CCGACGCGCGGACGCTTACCTCATCAAGCACGTGTCTGACGACTGGACGACGACGACG	43635	
QY	212	CCTCGCGCGCGCAGTGTCTGCGCGCGCGCGTCAACCGCGCAGCAGCCCAACCATGGACC	271	
Db	43636	TGTTAGCATCTCTCGCGCGGATCGGCACCGCATGAAGCGCGACTCCCGGCTCTCTGTCA	43695	
QY	272	TGGACGAGGAATTCGAGCTGTCTCGCACCGACCCCAAGCTGACGAGACATGAAGCGGC	331	
Db	43696	TCGACAACCTATCGACGAGCGCGCGCGCATCGACGCTCTTCGTCGACCTGCTGTGC	43755	
QY	332	GCATCATCGCCCTAATCTCTGAGCGCGCTGACGGCGACAAGCGCGCGCGATCGAGGA	389	
Db	43756	TGCTCTCTCTCGCGCGCGCGCAACGCTCGGAGAGCGAATTCGCGCGCTGCTGTGGAA	43813	
RESULT 8				
AAA58471 standard; DNA; 58857 BP.				
XX	AC	AAA58471;		
XX	31-OCT-2000	(first entry)		
XX	Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.			
KW	BLM gene cluster; bleomycin gene cluster; polyketide metabolite;			
KW	bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;			
KW	thiazoline; bithiazoline; microbial metabolite; sugar; ss.			
OS	Streptomyces verticillus.			
XX	Key	Location/Qualifiers		
FF	CDS	223..564		
FT		/*tag= a		
FT		/transl_except= (pos: 1..3, aa: Met)		
FT		/note= "ORF 30; encodes AAB07556"		
FT	CDS	561..2309		
FT		/*tag= b		
FT		/transl_except= (pos: 1..3, aa: Met)		
FT		/note= "ORF 29; encodes AAB07557"		
FT		2767..3486		
FT		/*tag= c		
FT		/note= "ORF 28; encodes AAB07558"		
FT	CDS	3527..5593		
FT		/*tag= d		
FT		/transl_except= (pos: 1..3, aa: Met)		
FT		/note= "ORF 27; encodes AAB07559"		
FT	CDS	5806..12294		
FT		/*tag= e		
FT		/note= "ORF 26; encodes AAB07560"		
FT	CDS	12291..15491		
FT		/*tag= f		
FT		/note= "ORF 25; encodes AAB07561"		
FT	CDS	15488..21013		
FT		/*tag= g		
FT		/note= "ORF 24; encodes AAB07562"		
FT	CDS	21010..24666		
FT		/*tag= h		
FT		/transl_except= (pos: 1..3, aa: Met)		
FT		/note= "ORF 23; encodes AAB07563"		
FT	CDS	24663..32690		
FT		/*tag= i		
FT		/note= "ORF 22; encodes AAB07564"		
FT	TT			

PT chemically modifying biological molecules.
XX
PS Claim 8; Page 97-136; 162pp; English.
XX
CC The present sequence represents the BLM (Bleomycin) gene cluster,
CC containing open reading frames (ORFs) 8-30. The proteins encoded by the
CC gene cluster are useful for producing peptides and/or polypeptide
CC metabolites, especially bleomycin or bleomycin analogues. They are also
CC useful for chemically modifying biological molecules to produce branched
CC methyl groups, and for coupling amino acids and fatty acids. They may be
CC reacted with an apo-carrier protein and coenzyme A to produce a holo-
CC carrier protein. The BLM gene cluster or catalytic domains can be used
CC individually or collectively to produce thiazolidine, thiazoline,
CC bithiazoline and bithiazoline-containing microbial metabolites. The BLM
CC gene cluster may also be used to produce sugars
XX
SQ Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;
Query Match 14.2%; Score 60.6; DB 3; Length 58857;
Best Local Similarity 49.5%; Pred. No. 0.041; Mismatches 159; Indels 0; Gaps 0;
Matches 156; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 91 GCCAGCGCATCCAGAAGACCGGGCCACCGTCGGCCGGTGGAGGACGGCAAGAACCGG 150
DB 8755 GCGCTGCCAAGCGCACGCCCGCGCGGACGGCGGGTACGTGCGCGCGGCACC 8814
QY 151 CCGGACGCGGAGACTCTGTTCCCGCGTCCGCGCCAGAGTCTCGGCTCGACTCGACGAA 210
DB 8815 GGCACCGAGAGATCTCGCGGCCACCGTCGCGAAGGTCTGGGCGTGGAGCGGTGCGC 8874
QY 211 GCCTCTGCGCGCGAGGTCTGCGCGCGCGCTCACCCGCCAGCGACCCCAACCATGGAC 270
DB 8875 ATCGAGCAACTACTTCTGTTGGCGCGACTCATCCGCGAGGTATGTTGCCACG 8934
QY 271 CTGAGCGAGGAATCGAGTGTTCGCGACCGACCCCAAGCTCGACGAGGACATGAGCGG 330
DB 8935 CGGGCCAGCGCGGGGTGAGGTACCGTGGCGGACCTGCACCGCGCACCCACCGTC 8994
QY 331 CGCATCATCGCTTAATCTGAGCGCGCTGAGCGCGACGCAAGCGCGCGCGATCGAGAA 390
DB 8995 CGGGCTGCGCGCGACCTGAGCGCGCGAGGACCTGCCCGGAGCGCGCGCTCACCGAA 9054
QY 391 ACCAAGCGGCTCATC 405
DB 9055 CCTTCGCGCTGATC 9069
RESULT 9
ABZ66792
ID ABZ66792 standard; DNA; 972 BP.
XX
AC ABZ66792;
XX
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 246.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Streptomyces mobaraensis.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
XX
PR 30-MAR-2001; 2001US-0279709P.
XX
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
Farnet CM, Zazopoulos E, Staffa A;
WPI; 2003-058435/05.
P-PSDB; ABP99329.
XX
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.
XX
PS Claim 29; Page 397; 511pp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism
XX
SQ Sequence 972 BP; 145 A; 356 C; 355 G; 116 T; 0 U; 0 Other;
Query Match 13.3%; Score 56.8; DB 7; Length 972;
Best Local Similarity 53.6%; Pred. No. 0.27; Mismatches 118; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 114 GGCCACCGTCGGCGGTGGAGGACGGCAAGAACCGGCCCGGACCGGACCTCGTTGC 173
DB 681 GTCCCTGTGGCGGACCGGATCGTGTGACCTTGGCCCCCGCGCGACCTCGCGG 740
QY 174 CCGCGTCGCCGAGGTCTCGGCTCGACCTCGACGAAAGCCCTCGCCCGCGAGGTCTGCG 233
DB 741 CGCCGCGCGAGGTCTCGGCTCGACGACACCGAGCCGAGGTGACCGCACACCG 800
QY 234 CCGCGGTACCCCGCGGACCGGACCCCAACCATGACCTGACGAGGAAATCGAGCTGGT 293
DB 801 CCGGTGACGCGCGGTGACGCGGACCGGCTGGGGGACCTCTGTGAGATCTGTGGGACGCT 860
QY 294 CCGCACCGACCCCAAGCTGGAGGACATGAAGCGCGC 333
DB 861 GCGGACCGCGGCCACGAGGTGGCGGACGTGCTGCTGCGC 900
RESULT 10
ABZ66808
ID ABZ66808 standard; DNA; 45055 BP.
XX
AC ABZ66808;
XX
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic gene cluster SEQ ID NO 277.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Streptomyces mobaraensis.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
XX
PR 30-MAR-2001; 2001US-0279709P.
XX
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.

Db 1039 CGCCGTGAGGTGGTGAACACCGACCGCGCGGTGCGCGGTGATCAGCATGCC 980
QY 303 CCCCAAGCTGACGAGGACATGAACGCGGCATCATCCCTTAATCTTGGAGCCCGTGA 362
Db 979 CGAATACGTGACGTGATCTCCCGCGCGGCAAGGCGCTGATCGAGCGCATCAGCG 920
QY 363 GCGGACAAAGCGCGCGGTGATCGAGAAACCAAGCGGTCTATCGACCTGTC 414
Db 919 CGAGCCAAAGTGGCGGTGATCAAGCACTGGACGGCATCTGCCACGCTTAC 868

RESULT 5

US-09-252-991A-5605
; Sequence 5605, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5605
; LENGTH: 2118
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5605

Query Match 12.8%; Score 54.4; DB 4; Length 2118;
Best Local Similarity 49.1%; Pred. No. 0.013;
Matches 173; Conservative 0; Mismatches 176; Indels 3; Gaps 1;
QY 63 GCGCGCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAGGACCGGCGCCACCGT 122
Db 565 GGGGTGTCGGGATCATCTACGATCGCGCGGCGAAGCTGACATCGACGCGCCAGCT 624
QY 123 CGGCGGTGGAGGAGCGGCAAGAACCGGCGGACGAGCGGACCTGTTGCCCGGTGCG 182
Db 625 GTGCTGAAGTCCGCAAGCGACCATCTGCGCGCGGCTCCGAGGCGATCCATCCAA 684
QY 183 CCAGTGTCTGCGCTCGACTCGAAGAGCCCTTCGCGCGCGAGGTCTGCGCCCGGGGT 242
Db 685 CCAGGCGATGCCCGTGTGATCATGACGAGGCGCTTGGCGGAGCGCGCTG---CCGCGCG 741
QY 243 CACCCCGCAGCGACCCCAACCATGACCTGACAGAGGAAATCGAGCTGGTCCGACCGA 302
Db 742 CGCGGTGACAGTGGTGGAAACACCCAGCGCGCGCGGTGCGCGCTGATCAGCATGCC 801
QY 303 CCCCAAGCTGAGGAGGACATGAAGCGGCGCATCTCGCCCTAATCTGAGGCGGTGA 362
Db 802 CGAATACGTGACGTGATCTGCCGCGCGGCGCAAGGCGCTGATCGAGCGCATAGCGG 861
QY 363 GCGGCAAAAGCGCGCGCGCATCGAGGAAACCAAGCGCGCTCATCGACCTGTC 414
Db 862 CGAGGCCAAGTCCCGGTGATCAAGCACCTGAGCGGCACTGCGCAAGTCTAC 913

RESULT 6

US-09-252-991A-3645
; Sequence 3645, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3645
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3645

Query Match 12.2%; Score 52; DB 4; Length 939;
Best Local Similarity 47.3%; Pred. No. 0.039;
Matches 194; Conservative 0; Mismatches 210; Indels 6; Gaps 1;

QY 14 CGGGGTGGCGCGGCACATGGCGCGCATACCTCACCAGCCCGCGAGCGCGCGGAC 73
Db 348 CGTTCGCCGAGCGCGCGTCTGAGCGCTTACGTGGCGACATCGACCGTTTGTGGCA 407
QY 74 TGACCAAGAGCGAGTTTGGCCAGCGCATCCAGAAAGACCGGCGCACCGTTCGGCGGTGG 133
Db 408 TCCTCCAGGCCAAGCGCGCGCTCGCAACTGCTCAGCAGGTGCGAGGGCTGTGCC 467
QY 134 AGAGCGCAAGAACCGCGCGCAGACGCGGACCTGTTGCCCGGTGCGCGAGTGTCTG 193
Db 468 TGTTCCTCAGCCTGCTGTGTGTGTTGTCACCCCTTACGACCTCAGTATCAGTCTATCG 527
QY 194 GCCTCGACCTCGACGAAGCCCTCGCGCGCAGGTCTGCGCGCGCGGTCAACCCGCGAG 253
Db 528 GCGCGTGGCGGAACTCACCAGCACCGCGCGCTGGCGCGCGGCGAGCTGGAGCGCG 587
QY 254 CGACCCCAACCATGACCTCGAGCGAGGAAATCGAGTGTGTCGCGACCGACCCCAAGTGG 313
Db 588 GGGTCACCTACAGCGCGGAGGACCACTGGCGCAACTGGGCGAGCGGCTTCAACCAATGG 647
QY 314 ACAGGACATACAGCGCGCATCATCGCCCTTAATCTTGGAGCGCGGTGAGCGGACAAAG 373
Db 648 CCGCGAGCTGAAGAG-----CATCTACGCGGACCTGGAAGACCGCTGGAGACAGA 701
QY 374 CGGCGCGGATCGAGGAAACCAAGCGGTCTATCGACCTGTTTCCGCGCGAGC 423
Db 702 CCGCGCGCTCTCGACAGCACCCAGCGCTCGAAGCTGCTTACGCGCAGC 751

RESULT 7

US-09-252-991A-3697
; Sequence 3697, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3697
; LENGTH: 1881
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3697

Query Match 12.2%; Score 52; DB 4; Length 1881;
Best Local Similarity 47.3%; Pred. No. 0.039;
Matches 194; Conservative 0; Mismatches 210; Indels 6; Gaps 1;

QY 14 CGGGCTGGCGCGGCACATGGCGCGCATACCTCACCAGCCCGGAGCGCGCGGAC 73

QY 74 TGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAAGACCGGCGCACCGTTCGGCGGTGGG 133
Db 26009 TCGTCATCGAGAGAGCGCGGAGCGCCACCGCGCCCGAACTGCTCGCGCCCGAAACCGCGCG 26068
QY 134 AGGACGGCAAGAACCGGCGCGACGACGCGGACCTCGTTGCCCGGTCCCGAGGTGCTCG 193
Db 26069 CCGAGCGGACGCTGCTGGTCCGAGGAGTGTGGCACGAGTGACCGTGGCCCTGATGATGT 26128
QY 194 GCGTCGACCTCGAGAGCCCTCGCGCGCGCGGCGGAGGTCTGGCGCCCGCGGTACCGCGCCAG 253
Db 26129 CCGCGCACAAAGCGCGCCCTGCGCGACCGAGCGCGCGCGCGCTGCGCGCGCGACCTGCTCG 26188
QY 254 CGACCCCAACCATGACCTGGAAGGAAATCGAGTGGTCCGACCGACCGACCGCAAGCTGG 313
Db 26189 CCCACCCGAGCTGACCGCGCGCGAGCTGGGTATACCCCTCATCACCGCGCACCGCGT 26248
QY 314 ACAGGA 320
Db 26249 TCGAGCA 26255

RESULT 10
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Bargett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostek, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL, 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1
Query Match
Best Local Similarity 12.0%; Score 51; DB 2; Length 44377;
Best Local Similarity 47.9%; Pred. No. 0.065;
Matches 147; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 14 CGGCGCTGGGCGCGGCGACATGCGGCGCATACGTCCTCAGCGCGCGCGAGCGCGCGAC 73
Db 25949 CCGAGGGGCGCGCGCGCGGCGCATCTCTCTGTTGCGCATCAGCGGACCAACGCGCAC 26008
QY 74 TGACCAAGAGCGAGTTGGCGCGCGCATCCAGAAAGACCGGCGCACCGTTCGGCGGTGGG 133
Db 26009 TCGTCATCGAGAGCGCGCGGAGCGCCACCGCGCCCGAACTGCTCGCGCCCGAAACCGCGCG 26068
QY 134 AGGACGGCAAGAACCGGCGCGCGACGACGCGGACCTCGTTGCCCGGTCCCGAGGTGCTCG 193
Db 26069 CCGAGCGGCGAGCTGCTGGTCCGAGGAGTGTGGCACGAGGTGACCGTCCCTGATGATGT 26128
QY 194 GCCTCGACCTCGAGAGCGCTCGCGCGCGAGGTCTGGCGCGCGCGGTTCACCGCGCGAG 253
Db 26129 CCGCGCACAAAGAGCGCGCCCTGCGCGACCGAGCGCGCGCGCTGCGCGCGCGACCTGCTCG 26188
QY 254 CGACCCCAACCATGAGACCTGGAAGGAAATCGAGTGGTCCGACCGACCGACCGCAAGCTGG 313
Db 26189 CCCACCCGAGCTGACCGCGCGCGAGCTGGGTATACCCCTCATCACCGCGCACCGCGT 26248
QY 314 ACAGGA 320
Db 26249 TCGAGCA 26255

RESULT 11
US-09-266-965-22
; Sequence 22, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09-03-12
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1224
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-22

Query Match
Best Local Similarity 11.8%; Score 50.4; DB 4; Length 1224;
Best Local Similarity 47.5%; Pred. No. 0.08;
Matches 150; Conservative 0; Mismatches 166; Indels 0; Gaps 0;
QY 99 CATCCAGAGACCGGCGCACCGTTCGGCGGTGGAGACCGCAAGAACCGCGCGCGACGA 158
Db 861 CGTCCAGCAGGCGCTGCGCGCGCGGTCAACCGGAGATGAGATCGCGCGCGACAGGT 920
QY 159 CGCGACCTCGTTGCCCGCGTGGCCCGGTGCGCCAGGTCTCGACCTCGACGAAGCCCTCGC 218

104 AGAAGGACCGGGCCACCGTCGGCGGTGGAGGACGGCAAGAACCGGCCCGACGACGCGG 163
437 GCCTGCGCAAGGGCGAGTGTGTACGGGTCCAGGCCGCGAGATGATTCCCGGCGACGCGG 496
164 ACCTCGTTGCCCGGTGCGCCAGGTGCTCGGCTCGACCT---CGAGAAAGCCCTCGCCG 220
497 AAGTCATCGAAGGGGTGGCGCGTCAACAGAGGCGCCATCACCGGCGAATCGGCGCGG 556
221 CCGCAGGTCTGCGCCCGGGGTCAACCCCGCAGCGACCCCAACCATGGACCTGGACGAGG 280
557 TGATCCCGAGTCCGGCGGGGATCGCTCGGGGTGACCGGCAACACCCAGTGTGTCCG 616
281 AAATCGAGCTGTCCGACCGACCCCAAGTGGACGAGGACATGAAGCGGCGCATCATCG 340
617 ACTGGCTGCTGTACGATCGGCGCCCAACCCGGCGAGTCGACCTGGACCGGATGATCG 676
341 CCCTAATCTTGGAGCGCGTGAAGCGGACAGCGGCGCGATCGAGGAAACCAAGCGGC 400
677 CCCTGGTGAAGGCGCCAAAGCGGCGAAGACCCCAAGAGTGGCGCTGGACATCTCTGC 736
401 TCATCGACCTGTTC 415
737 TGATCGGCTTGACCC 751

Search completed: August 6, 2004, 13:42:31
Job time : 58.2032 secs

QY 104 AGAAGGACCGGGCCACCGTCGGCGGTGGAGGACGGCAAGAACCGGCCCGACGACGCGG 163
Db
QY 437 GCCTGCGCAAGGGCGAGTGTGTACGGGTCCAGGCCGCGAGATGATTCCCGGCGACGCGG 496
Db
QY 164 ACCTCGTTGCCCGGTGCGCCAGGTGCTCGGCTCGACCT---CGAGAAAGCCCTCGCCG 220
Db
QY 497 AAGTCATCGAAGGGGTGGCGCGTCAACAGAGGCGCCATCACCGGCGAATCGGCGCGG 556
QY 221 CCGCAGGTCTGCGCCCGGGGTCAACCCCGCAGCGACCCCAACCATGGACCTGGACGAGG 280
Db
QY 557 TGATCCCGAGTCCGGCGGGGATCGCTCGGGGTGACCGGCAACACCCAGTGTGTCCG 616
QY 281 AAATCGAGCTGTCCGACCGACCCCAAGTGGACGAGGACATGAAGCGGCGCATCATCG 340
Db
QY 617 ACTGGCTGCTGTACGATCGGCGCCCAACCCGGCGAGTCGACCTGGACCGGATGATCG 676
QY 341 CCCTAATCTTGGAGCGCGTGAAGCGGACAGCGGCGCGATCGAGGAAACCAAGCGGC 400
Db
QY 677 CCCTGGTGAAGGCGCCAAAGCGGCGAAGACCCCAAGAGTGGCGCTGGACATCTCTGC 736
QY 401 TCATCGACCTGTTC 415
Db 737 TGATCGGCTTGACCC 751

RESULT 15
US-09-252-991A-9429
; Sequence 9429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9429
; LENGTH: 2112
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9429

Query Match 11.8%; Score 50.2; DB 4; Length 1344;
Best Local Similarity 47.7%; Pred. No. 0.087;
Matches 179; Conservative 0; Mismatches 193; Indels 3; Gaps 1;
QY 44 ACCTCTCTACCGCGCGGCGGCGGAGTGTGCGGCTCGACCT---CGACGAAGCCCTCGCGG 103
Db 1015 AGGGCTCTTCGCCAATAGCGGCGCGCGACGCGAGTTTCGAGAGCGTCTGCTGCCAGCA 956
QY 104 AGAAGACCGGCGCCCGTGGCGGAGGAGCGCAAGAACCGGCCCGCGAGCGCGG 163
Db 955 GCGTCGCAAGGCGGACGTGTACGGGTCCAGGCGCGGAGATGATTCCCGCGGACGCGG 896
QY 164 ACCTCGTTCCCGCGCGCGGCGGAGTGTGCGGCTCGACCT---CGACGAAGCCCTCGCGG 220
Db 895 AAGTCATCGAAGGGTGGCGGCGGTCAACAGGCGGCCATCACCGGCGAATCGGCGCGG 836
QY 221 CCGCAGGTCTGCGCCCGCGCGTCAACCCCGCAGCGACCCCAACCATGGACCTGGACGAGG 280
Db 835 TGATCGCGAGTCCGCGCGCGATCGTCTGGCGGTGACCGGCAACACCCAGGTGTGTCCG 776
QY 281 AAATCGAGCTGTCCGACCGACCCCAAGTGGACGAGGACATGAAGCGGCGCATCATCG 340
Db 775 ACTGGTCTGTGTACGATCGGCGCCCAACCCGGCGAGTCGACCTGGACCGGATGATCG 716
QY 341 CCCTAATCTTGGAGCGCGTGAAGCGGACAGCGGCGCGATCGAGGAAACCAAGCGGC 400
Db 715 CCCTGGTGAAGGCGCCAAAGCGGCGAAGACCCCAAGAGTGGCGCTGGACATCTCTGC 656
QY 401 TCATCGACCTGTTC 415
Db 655 TGATCGGCTTGACCC 641

Query Match 11.8%; Score 50.2; DB 4; Length 2112;
Best Local Similarity 47.7%; Pred. No. 0.088;
Matches 179; Conservative 0; Mismatches 193; Indels 3; Gaps 1;
QY 44 ACCTCTCTACCGCGCGGCGGCGGAGTGTGCGGCTCGACCT---CGACGAAGCCCTCGCGG 103
Db 377 AGGGCTCTTCGCCAATAGCGGCGCGGAGTGTGCGGCGGAGTGTGCGGCGGAGTGTGCGG 436

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:22:34 ; Search time 1649.94 Seconds
(without alignments)
7710.163 Million cell updates/sec

Title: US-09-855-340A-2
Perfect score: 426
Sequence: 1 atgcgcacacacccgggct.....acctgtccgcggagctga 426

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	69.8	16.4	935	29	CNS006XK
C 2	63	14.8	925	29	CNS0091P
C 3	62.8	14.7	1101	29	CNS017SY
C 4	62.4	14.6	582	10	BE401996

5	62.4	14.6	582	13	BQ607313	BQ607313
6	62.4	14.6	594	14	CA597375	CA597375
7	62.4	14.6	620	10	BE417374	BE417374
8	62.4	14.6	673	13	BQ839181	BQ839181
9	62.4	14.6	690	14	CD869706	CD869706
10	62.4	14.6	694	14	CD865484	AZ02.112G
11	62.4	14.6	872	10	BE415929	MUG002.E0
12	62.4	14.6	1035	14	CK212590	CK212590
13	62.2	14.6	599	13	BQ470159	BQ470159
14	62	14.6	553	14	CA497279	WHE32255 C
15	61.6	14.5	424	10	BE400165	BE400165
16	61.2	14.4	479	13	CA025140	HZ51F07r
17	61.2	14.4	925	29	CNS0091P	AL053013 Drosophil
18	60.8	14.3	474	10	BE498929	WHE0968 C
19	60.8	14.3	486	10	BE423849	BE423849
20	60.8	14.3	590	14	CD934503	OV.CO1E14
21	60.8	14.3	601	12	EM140310	WHE0474 g
22	60.8	14.3	611	10	BE401863	BE401863
23	60.8	14.3	611	13	BQ607482	BRY 3376
24	60.8	14.3	633	14	CA731367	wipic.pk0
25	60.8	14.3	640	10	BE427136	BE427136
26	60.8	14.3	819	13	BQ807131	BQ807131
27	60.8	14.3	1049	14	CK184005	FGAS01665
28	60.6	14.2	467	13	BU979539	HA16HU9r
29	60.6	14.2	472	13	CA024821	HZ50G11r
30	60.6	14.2	491	13	BU980307	BU980307
31	60.6	14.2	511	13	BU981850	HA24N03r
32	60.6	14.2	537	14	CB883553	H002J16w
33	60.6	14.2	550	13	BQ470861	BQ470861
34	60.6	14.2	569	13	BU997340	BU997340
35	60.6	14.2	1026	12	BG343655	HVSMG000
36	60.2	14.1	474	10	BE416246	BE416246
37	60.2	14.1	520	10	BE417736	MUG024.A0
38	60.2	14.1	619	13	BQ460128	HA07C13r
39	60	14.1	361	10	BE425095	WHE0080.A
40	59.6	14.0	1201	13	EX360624	EX360624
41	59.2	13.9	1123	14	CB218083	CB218083
42	59	13.8	440	14	CB883338	HQ01021w
43	59	13.8	627	13	BU999317	BU999317
44	59	13.8	748	29	CG271729	CG3AV22TV
45	58.6	13.8	1097	12	BI952002	HVSMEM000

ALIGNMENTS

RESULT 1
CNS006XK/c 935 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BAC14N09 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066051
VERSION AL066051.1 GI:4945019
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 935)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>
Melanogaster BAC library was prepared by Kazutoyo Oseogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 935
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPI-98"
/note="end : T7"

ORIGIN
Query Match 16.4%; Score 69.8; DB 29; Length 935;
Best Local Similarity 30.8%; Pred. No. 1.8;
Matches 116; Conservative 93; Mismatches 168; Indels 0; Gaps 0;

Qy 3 GCGCAACACACCGGGCTGGGGCGCGGCACATGGCGCGCATAGTCTCACCGCCGCGA 62
Db 920 GSGCGGCGGSGGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 861
Qy 63 GCGCGCGCGGACTACCAAGAGCGAGTGGCCAGCGCGCATCCAGAGGACCGGGCCACCGT 122
Db 860 GSSCCG 801
Qy 123 CGCGCGGTGGAGAGCGCAAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
Db 800 CG 741
Qy 183 CGAGGTCTCGGCTCGACCTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 242
Db 740 GSGGCG 681
Qy 243 CACCCCG 302
Db 680 CG 621
Qy 303 CCGCAGCTGAGAGGACATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
Db 620 CG 561
Qy 363 GCGCGCAACGCGCGCG 379
Db 560 AGSGKMSAGSGRCGG 544

RESULT 2
CNS0091P/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL053013
AL053013.1 GI:4934461
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 925
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPI-98"
/note="end : TET3"

ORIGIN
Query Match 14.8%; Score 63; DB 29; Length 925;
Best Local Similarity 15.1%; Pred. No. 16;
Matches 53; Conservative 180; Mismatches 114; Indels 4; Gaps 2;

Qy 3 GCGCAACACACCGGGCTGGGGCGCGGCACATGGCGCGCATAGTCTCACCGCCGCGA 62
Db 900 SNSBSCSSSSBSSTSSSSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBS 841
Qy 63 GCGCGCGGACTACCAAGAGCGAGTGGCCAGCGCGCATCCAGAGGACCGGGCCACCGT 122
Db 840 EECWCSSSSSCCGAGARGVKVRASGAGVRGGSGGASASHSSSACBSSSSSCSASOW 701
Qy 123 CGCGCGGTGGAGGACGCGCAAGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
Db 780 SASSSSSASSSSSRSGGAGSGGASSSSSSSSSSSSSSSSSSSSSSSSSSSSS--VSC 723
Qy 183 CGAGTCTCGGCTCGACCTCGACGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 242
Db 722 SSVASSMSCSSBS 663
Qy 243 CACCCCG 302
Db 662 SSCSSSSMSASSSSSASSSSSSSSSSSSSSSSSSSSS--SACGBSMSSGGSGSVASGMSVV 605
Qy 303 CCCCAAGCTGAGGACATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
Db 604 SSSGGRSSGGGGGSGGSGSSSSSSSSSSSSSSSSSSSSVSCSCSCGCMCRCSA 554

RESULT 3
CNS017SY
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence Sp6 end of BAC BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridae; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a

collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC
 library (Bros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES
 source Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN37L08"
 /clone_lib="DrosBAC"
 /plasmid="pBelOBAC11"
 /note="end : SP6"

ORIGIN

Query Match 14.7%; Score 62.8; DB 29; Length 1101;
 Best Local Similarity 16.5%; Pred. No. 17;
 Matches 69; Conservative 184; Mismatches 165; Indels 0; Gaps 0;
 QY 9 CACACGGGGCTGGGGCGGCGACATGGCCCATACGTCCTCACCAGCGGCGGCGGCGGCG 68
 DB 632 CCCACACGGGACGCGRRKCAACKAGMSGCCGSGSGSCCGKKKXGVRGVRVCCA 691
 QY 69 CGGACTGACCAAGAGCGAGCTTGGCCAGGCGCATCCAGAGGACCGGCGCACCGTCGCGCG 128
 DB 692 GGGASCACAAACGGCCAKVACCCSSSSASSSSSCASTSSASRGVMVSSCACSGSGG 751
 QY 129 GTGGAGGACCGCAGAACCGGCGGACGACCGGACCTCTGTTGCCGCGTCGCCAGGT 188
 DB 752 AGACGAGGAGGGGSGGGGCGCAGGVCAGAACSSASCSASMGVSSGSSCSASGSGCG 811
 QY 189 GCTCGGCTCGACCTCGAGGACCTCGCGCGCGAGCTCTGCGCGCGGCGTCAACCC 248
 DB 812 VSSCAVSSASSSVNSKVASAVSCSAVASGMSAGVSSSCRSSVASVSSAAASVSSV 871
 QY 249 GCGAGCGACCCCAACCATCGAGCTGAGAGGAAATCGAGCTGTGTCGACCGACCCCA 308
 DB 872 SSSSSSVVAAAVGASSSSASSASMAVAAAVSVSVSVSSSSSSSSSSSV 931
 QY 309 GTGACGAGGACATGAACGGCGGATCATCCCTTAATCTTGGAGCGCGTACGCGGA 368
 DB 932 VSAVAASASSVSSSSSSSSSSSSSVSSASVSSVAVSMASVSSSSSSSVSVVAVAA 991
 QY 369 CAAGGGCGGCGGATCGAGAAACCAAGGCGTCTCGACCTGTTCGCGCGGAGCTGA 426
 DB 992 SAAAAAAAAAASSSASAVAVSSSSSSSSSSSSSSSSSSSSSVSVSV 1049

RESULT 4
 BE401996
 LOCUS
 DEFINITION
 CSB003D02F990308 ITEC CSB Wheat Endosperm Library Triticum aestivum
 CDNA clone CSB003D02, mRNA sequence.

ACCESSION
 BE401996
 VERSION
 BE401996.1 GI:9361464

KEYWORDS
 SOURCE
 ORGANISM

Triticum aestivum (bread wheat)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.

REFERENCE
 AUTHORS
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,
 Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,
 Jouari, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,
 Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,
 Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.,
 International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae

JOURNAL
COMMENT

Unpublished (2000)
 Contact: Appels R
 Div. of Plant Industry, CSIRO
 Canberra ACT 2601 AUSTRALIA
 Tel: 61 62 465496
 Fax: 61 62 465000
 Email: rudi@pi.csiro.au
 International Triticeae EST Cooperative (ITEC)
<http://wheat.pw.usda.gov/genome>.
 Location/Qualifiers

FEATURES
source

1..582
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Wyuana"
 /db_xref="taxon:4565"
 /clone="CSB003D02"
 /tissue_type="endosperm"
 /dev_stage="8-12 days post anthesis"
 /lab_host="Escherichia coli SOLR"
 /clone_lib="ITEC CSB Wheat Endosperm Library"
 /note="Vector: Lambda Zap/Bluescript; Site 1: XhoI;
 Site 2: EcoRI; plants grown in phytotron with 18C/13C
 (day/night) 16 hour light. M13 Reverse sequencing primer
 used. 1.0 Kbp average insert size."

ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 582;
 Best Local Similarity 50.9%; Pred. No. 19;
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
 QY 67 GCGGACTGACCAAGACCGAGTGGCCAGGCGGATCCAGAGG---ACCGGGCCACCGTC 123
 DB 78 GCGGAGAGAGAGAGAGGCGGTGACCGGTCCACCAAGCGCGCTCCAGTTCCCCGTC 137
 QY 124 GCGCGGTGGAGGACGCAAGAACCGGCGCGACCGGACCTGTTGCCGCGTCGCC 183
 DB 138 GCGCGCATCGGCGCTTCTCAGAAGGCGCGTACGCGAGCGGCTCGCTCGGCGCC 197
 QY 184 CAGTGTCTGCGCTTCGACCTCGACGAAGCCCTCGCGCGCGAGGTCTGCGCGCGCGGTC 243
 DB 198 CCGTCTACCTCGCGCGCGTCTCTAGTACTCTCGCGCGCGAGGTCTGAGAGTCTCGCGCG 257
 QY 244 ACCCGCGCAGGACCCCAACCATGGACCTGGAGGAGAAATCGAGCTGTTCCGACCGAC 303
 DB 258 AACCGCGCAAGGACAACAAGAGACCCCGCATGTGCGCGCGCACCTGCTCTCGGCATC 317
 QY 304 CCCAAGTGGACGAGGACATGAAGCGCGCATCATGCCCTTAATCTTGGAGCGCGGTGAG 363
 DB 318 CGCAACGACGAGGAGCTCGGCAAGCTCTCGCGCGGTACCATCGGCGACGCGCGGTG 377
 QY 364 CGCAGCAGGCGGCGGATCGAGGAAACCAAGCGGCTCATCGA 407
 DB 378 CTCCCCAACATCAACCCCGTCTCTCCCAAGAGGCGCTCGA 421

RESULT 5

BE407313

LOCUS

DEFINITION

BRY 3207 wheat EST endosperm library Triticum aestivum CDNA 5',

mRNA sequence.

ACCESSION

BE407313

VERSION

BE407313.1 GI:21556642

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

REFERENCE

AUTHORS

Clarke, B., Lambrecht, M. and Rhee, S.Y.

Arabidopsis genomic information for interpreting wheat EST

SEQUENCES

FUNCTION

Integr. Genomics 3 (1-2), 33-38 (2003)

MEDLINE

PUBMED

COMMENT

22478026

Contact: Lambrecht M

The Arabidopsis Information Resource

Carnegie Institution of Washington, Dept. of Plant Biology

260 Panama Street, Stanford, CA 94305, USA

Tel: 1 650 325 1521 x 251

Fax: 1 650 325 3748

Email: rhes@acoma.stanford.edu.

Location/Qualifiers

1. 582

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Wyuana"

/db_xref="taxon:4565"

/tissue_type="endosperm"

/dev_stage="developing endosperm tissue 8, 10 and 12 DPA

(days post anthesis)"

/clone_lib="wheat EST endosperm library"

FEATURES

source

ORIGIN

Query Match 14.6%; Score 62.4; DB 13; Length 582;

Best Local Similarity 50.9%; Pred. No. 19;

Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGCGCGATCCAGAAGG---ACCGGCCACCGTC 123

Db 78 GCGGAGAGAGGAAGAAGCGGTGACCCGGTCCACCAAGCGCGCTCCAGTTCCCGCTC 137

QY 124 GCGCGTGGGAGGACGCGAAGACCGCGCCGAGCGAGCGAGCTGTTGCCCGCGTCGCC 183

Db 138 GCGCGATCGGGCTTCTCAAGAGCGCGGTACCGCGAGCGGTGCGTCCGGCGCC 197

QY 184 CAGTGTCTGGCTCGACCTCGACGAGCCCTGCGCGCGAGGTGTCGCCCGCGCGTC 243

Db 198 CCGTCTACTCGCCCGCTCTCTGAGTACCTCGCGCGAGGTGTCGAGCTCGCGCGC 257

QY 244 ACCCGCGAGCGACCCCAACCATGGACCTGGACGAGGAATCGAGTGTTCGCGACCGAC 303

Db 258 ACGCGCGCAAGGACAAAGAACCCGATCGTCCGCGCACCTGCTCTCGCCATC 317

QY 304 CCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTGTGAGCGCGTGAG 363

Db 318 CGCAACGACGAGGCTCGCAAGTCTGCGCGGTCAACATCGCGACGCGCGCGTG 377

QY 364 CGGCAAGCGCGCGCGATCGAGGAACCAAGCGCTCATCGA 407

Db 378 CTCCCAACATCAACCCCGTGTCTCTCCCAAGAGGCGCCCTCGA 421

RESULT 6

CA597375

LOCUS

DEFINITION wpalc.pk016.n10 wpalc Triticum aestivum cDNA clone wpalc.pk016.n10

5' end, mRNA sequence. EST 21-NOV-2002

ACCESSION CA597375.1 GI:25150171

VERSION CA597375

KEYWORDS EST

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

1 (bases 1 to 594)

Tingey, S.V., Moore, G., Griffiths, S., Powell, W., Wolters, P.,

Dolan, M., Hainey, C., Miao, G., Caraher, N. and Hanafey, M.K.

DuPont Wheat cDNA Sequence in collaboration with the John Innes

Center 1

Unpublished (2002)

Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

Location/Qualifiers

1. 594

/organism="Triticum aestivum"

/mol_type="mRNA"

/db_xref="taxon:4565"

/clone_lib="wpalc.pk016.n10"

/tissue_type="anthers"

/lab_host="DH10B"

/clone_lib="wpalc"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 594;

Best Local Similarity 50.9%; Pred. No. 19;

Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCCGACTGACCAAGAGCGAGTTGGCCAGCGCGATCCAGAAGG---ACCGGCCACCGTC 123

Db 72 GCGGAGAGAGGAAGAAGCGGTGACCCGGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTC 131

QY 124 GCGCGTGGGAGGACGCGAAGACCGCGCCGAGCGAGCGAGCTGTTGCCCGCGTCGCC 183

Db 132 GCGCGCATCGGGCGCTTCTCAAGAGAGGCGCGCTACGCGAGCGCGTCGGCTCCGGCGCC 191

QY 184 CAGTGTCTGGCTCGACCTCGACGAGCCCTGCGCGCGAGGTGTCGCCCGCGCGTC 243

Db 192 CCGTCTACTCGCCCGCTCTCTGAGTACCTCGCGCGAGGTGTCGAGCTCGCGCGC 251

QY 244 ACCCGCGAGCGACCCCAACCATGGACCTGGACGAGGAATCGAGTGTTCGCGACCGAC 303

Db 252 AACCGCGCAAGGACAAAGAGAGCGCGCATCGTCCGCGCGCACCTGTCTCTCGCCATC 311

QY 304 CCCAAGCTGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTGTGAGCGCGTGAG 363

Db 312 CGCAACGACGAGGCTCGCAAGTCTGCGCGGTCAACATCGCGACGCGCGCGTG 371

QY 364 CGGCAAGCGCGCGCGATCGAGGAACCAAGCGCTCATCGA 407

Db 372 CTGCGCAACATCAACCCCGTGTCTCTCCCAAGAGGCGCCCTCGA 415

RESULT 7

BE417374

LOCUS

DEFINITION MUG020.D01R990620 ITEC MUG Wheat Spikelet Library Triticum aestivum

cDNA clone MUG020.D01, mRNA sequence. linear EST 24-JUL-2000

ACCESSION BE417374

VERSION BE417374.1 GI:9415220

KEYWORDS EST

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Triticum.

1 (bases 1 to 620)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,

Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,

Gustafson, P., Herrmann, R.G., Holtan, T., Jacquemin, J.M., Jia, J.,

Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,

Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,

Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.

International Triticeae EST Cooperative (ITEC): Production of

Expressed Sequence Tags for Species of the Triticeae

Unpublished (2000)

Contact: Ogihara Y

Kihara Institute for Biological Research, Yokohama City University

Maioaka-cho 641-12, Totsuka-ku, Yokohama 244-0813, JAPAN

Tel: 81 45 820 1903

Fax: 81 45 820 1901
 Email: ogihara@yokohama-cu.ac.jp
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.

FEATURES

Location/Qualifiers
 1..620
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Norin 26"
 /db_xref="taxon:4565"
 /clone="WUG020.D01"
 /tissue_type="young spikelets"
 /dev_stage="Feekes' scale 6-7"
 /clone_lib="ITEC MUG Wheat Spikelet Library"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; M13 Reverse sequencing primer used. 1.2 Kbp average insert size."

ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 620;
 Best Local Similarity 50.9%; Pred. No. 19;
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCGGACTACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAGG---ACCGGCCACCGTC 123
 Db 5 GCGGAGAGAGAGAGCGGTGACCGGTCCACCAAGCGCGCTCCAGTTCCCGTC 64

QY 124 GCGCGTGGAGAGCGGCAAGAACCGGCCGACGACGCGGACCTGTGCGCGCGTCGCC 183
 Db 65 GCGCGCATCGGCGCTTCTCCTCAAGAAGGCGGTACGCGACGCGGTGCGGTCCGCGCC 124

QY 184 CAGGTCTCGGCTGACCTGACGAGGACCTCGCGCGCGAGGTCGCGCGCGGTC 243
 Db 125 CCGGTCTACCTCGCGCGGCTCTCGAGTACCTCGCGCGGAGGTGCTGGAGTCCGCGG 184

QY 244 ACCCGCCACGACCCCAACCATGACCTGACGAGGAATCGAGTGTTCGCCACCGAC 303
 Db 185 AACCGCGCAAGACACCAAGAGACCCGACATGTCGCGCGCACCTGTCTCCGCCATC 244

QY 304 CCGAGCTGGAGAGACATGAAGCGCGCATCATCGCCCTAATCTCGAGGCGCGTGAG 363
 Db 245 CGCAACGACGAGAGCTCGGCAAGTGTCTCGCGCGGTCCACATCGCGACGCGCGGCG 304

QY 364 CGCGACAAGCGCGCGCATCGAGGAACCAAGCGGCTCATCGA 407
 Db 305 CTCCCCAACATCAACCCCGTGTCTCCCAAGAGGCGCTCGA 348

RESULT 8
 BQ839181 673 bp mRNA linear EST 08-AUG-2002
 LOCUS WHE4163_C06_E112S Wheat CS whole plant cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE4163_C06_E11, mRNA sequence.

ACCESSION BQ839181
 VERSION BQ839181.1 GI:22143503
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum

REFERENCE Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K.,
 Dvorak, J., Lazo, G.R., Rausch, C.J., Wilson, C. and Woo, J.

TITLE The structure and function of the expressed portion of the wheat
 genomes - Chinese Spring whole plant cDNA library

JOURNAL Unpublished (2002)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818

Email: andersn@pw.usda.gov
 Sequences have been trimmed
 quality sequence with phred score less than 20
 Seq primer: SK primer.

FEATURES

Location/Qualifiers
 1..673
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE4163_C06_E11"
 /tissue_type="Roots, leaves, crown, stem and sheath"
 /dev_stage="Adult"
 /lab_hosts="E. coli SOLR"
 /clone_lib="Wheat CS whole plant cDNA library"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid
 pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant
 tissues from wheat cv. CS grown to full tillering stage in
 greenhouse were collected at University of California,
 Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
 RNA was prepared from leaves (young leaf and third leaf),
 whole roots, crown, stem and sheath tissues, and then
 equal quantities of RNA were pooled from the these
 samples. PolyA was purified from the pooled RNA, a cDNA
 library was made, and the cDNA clones were in vivo excised
 to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.
 Akhunov, J. Dvorak) at the University of California,
 Davis. Colony plating, plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."

ORIGIN

Query Match 14.6%; Score 62.4; DB 13; Length 673;
 Best Local Similarity 50.9%; Pred. No. 19;
 Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;

QY 67 GCGGACTGACCAAGAGCGAGTTGGCCAGGCGCATCCAGAAGG---ACCGGCCACCGTC 123
 Db 65 GCGGAGAGAGAGAGGCGGTGACCGGTCCACCAAGCGCGCTCCAGTTCCCGTC 124

QY 124 GCGCGTGGAGAGCGGCAAGAACCGGCCGACGACGCGGACCTGTGCGCGCGTCGCC 183
 Db 125 GCGCGCATCGGCGCGCTTCTCCTCAAGAAGGCGGTACGCGCGCGGTGCTGGAGTCCGCGGC 184

QY 184 CAGGTGTCTCGGCTTGCACCTCGAGAAAGCCCTCGCGCGCGAGGTCTGCGCGCGCGTC 243
 Db 185 CCGGTCTACCTCGCGCGGCTCTCGAGTACCTCGCGCGCGAGGTGCTGGAGTCCGCGGC 244

QY 244 ACCCGCCACGACCCCAACCATGACCTGACGAGGAATCGAGTGTTCGCCACCGAC 303
 Db 245 AACCGCGCAAGACACCAAGAGACCCCGCATGTCGCGCGCGCTCTCTCGCCATC 304

QY 304 CCCAAGCTGGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTCGAGGCGCGTGAG 363
 Db 305 CGCAACGACGAGAGCTCGGCAAGTGTCTCGCGCGGTCCACATCGCGCACGCGCGGTG 364

QY 364 CGCGACAAGCGCGCGGATCGAGGAACCAAGCGGCTCATCGA 407
 Db 365 CTCCCCAACATCAACCCCGTGTCTCCCAAGAGGCGCTCGA 408

RESULT 9

CD869706
 LOCUS AZ02.112G03F001120 AZ02 Triticum aestivum cDNA clone AZ02112G03,
 DEFINITION mRNA sequence.

ACCESSION CD869706
 VERSION CD869706.1 GI:32553522
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum

REFERENCE Anderson, O.D., Akhunov, E., Chao, S., Crossman, C., Deal, K.,
 Dvorak, J., Lazo, G.R., Rausch, C.J., Wilson, C. and Woo, J.

TITLE The structure and function of the expressed portion of the wheat
 genomes - Chinese Spring whole plant cDNA library

JOURNAL Unpublished (2002)
 COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818

Poideae; Triticeae; Triticum.

REFERENCE

1 (bases 1 to 690)
Genoplatte.
Genoplatte, a major partnership french program in plant genomics
Unpublished (2003)
Contact: Genoplatte
Genoplatte
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)
and <http://genoplatte-info.infobiogen.fr>.

FEATURES

source
Location/Qualifiers
1..690
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="AZO2112603"
/tissue_type="root"
/clone_lib="AZO2"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 690;
Best Local Similarity 50.9%; Pred. No. 19;
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
Qy 67 GCCGACTGACCAAGAGCGAGTTGGCCAGGCGGATCCAGAGG---ACCGGCGCACCGTC 123
Db 85 GCGGAGAGAGAGAGAGAGCGGTGACCCGGTCCACCAAGCGCGCTCCAGTTCCCGCTC 144
Qy 124 GCGCGGTGGAGAGAGCGCAAGAACCGCGCGCGGACGAGCGGACCTGTTTCCCGCGTCGCG 183
Db 145 GCGCGCATCGGGCGCTTCTCAAGAGAGGCGGTACGCGAGCGGTGCTTCCCGCGCC 204
Qy 184 CAGTGCTCGGCTCGACCTCGACGAGAGCCCTCGCGCGCGAGGTGTGCGCCCGCGCGTC 243
Db 205 CCGGTCTACCTCGCGCGCGTCTCGAGTACCTCGCGCGCGAGGTGTGAGCTCGCGCG 264
Qy 244 ACCCGCGGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
Db 265 ACAGCGCGGAG 324
Qy 304 CCAAGCTGAGGAG 363
Db 325 CGCAACGAGGAG 384
Qy 364 CGGAGAGAGGCGGCGGCGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
Db 385 CTCCCAACATCAACCCCGTGTGCTCTCCCAAGAGGCGGCTCGA 428

RESULT 10

CD865484
LOCUS
DEFINITION
AZO2.101A18F010111 AZO2 Triticum aestivum cDNA clone AZO2101A18,
mRNA sequence.
CD865484
VERSION
CD865484.1 GI:32549300
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 694)

REFERENCE

Genoplatte.
Genoplatte, a major partnership french program in plant genomics
Unpublished (2003)
Contact: Genoplatte
Genoplatte
93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french
plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>)
and <http://genoplatte-info.infobiogen.fr>.

FEATURES

source
Location/Qualifiers
1..694
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="recital"
/db_xref="taxon:4565"
/clone="AZO2101A18"
/tissue_type="root"
/clone_lib="AZO2"

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 694;
Best Local Similarity 50.9%; Pred. No. 19;
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
Qy 67 GCCGACTGACCAAGAGCGAGTTGGCCAGGCGGATCCAGAGG---ACCGGCGCACCGTC 123
Db 76 GCGGAGAGAGAGAGAGAGCGGTGACCCGGTCCACCAAGCGCGCTCCAGTTCCCGCTC 135
Qy 124 GCGCGGTGGAGAGAGCGCAAGAACCGCGCGCGGACGAGCGGACCTGTTGCGCGCGTCGCG 183
Db 136 GCGCGCATCGGGCGCTTCTCAAGAGAGGCGGTACGCGAGCGGTGCTTCCCGCGCC 195
Qy 184 CAGTGCTCGGCTCGACCTCGACGAGAGCCCTCGCGCGCGAGGTGTGCGCCCGCGCGTC 243
Db 196 CCGGTCTACCTCGCGCGCGTCTCGAGTACCTCGCGCGCGAGGTGTGAGCTCGCGCG 255
Qy 244 ACCCGCGGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
Db 256 AACCGCGGAG 315
Qy 304 CCAAGCTGAGGAG 363
Db 316 CGCAACGAGGAG 375
Qy 364 CGGAGAGAGGCGGCGGCGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 407
Db 376 CTCCCAACATCAACCCCGTGTGCTCTCCCAAGAGGCGGCTCGA 419

RESULT 11

BE415929
LOCUS
DEFINITION
MUG002.E06R990520 ITEC MUG Wheat Spikelet Library Triticum aestivum
cDNA clone MUG002.E06, mRNA sequence.
BE415929
VERSION
BE415929.1 GI:9413775
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 872)

REFERENCE

Genoplatte.
Genoplatte, a major partnership french program in plant genomics
Unpublished (2000)
Contact: Ogiwara Y
Kihara Institute for Biological Research, Yokohama City University
Maoka-cho 641-12, Totsuka-Ku, Yokohama 244-0813, JAPAN
Tel: 81 45 820 1903
Fax: 81 45 820 1901

Email: ogihara@yokohama-cu.ac.jp
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers

FEATURES

1..872
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Norin 26"
/db_xref="taxon:4565"
/clone="MUG002.E06"
/tissue_type="young spikelets"
/dev_stages="Feekes' scale 6-7"
/clone_lib="ITEC MUG Wheat Spikelet Library"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; M13 Reverse sequencing primer used. 1.2 Kbp average insert size."

ORIGIN

Query Match 14.6%; Score 62.4; DB 10; Length 872;
Best Local Similarity 50.9%; Pred. No. 19;
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
QY 67 GCGGACTGACCAAGAGCGAGTTGCCAGGCGCATCCAGAGG---ACGGGCCACCGTC 123
Db 81 GCGCGAGAGAGGAAGCGGTGACCGGTCCTCAAGGCGCGCTCAAGTTCCTCCGTC 140
QY 124 GCGCGTGGGAGAGCGCAAGAACCGCGCGACGCGGACCTCGTTGCCCGCTGCC 183
Db 141 GCGCGCATCGGGGCTTCTCAAGAGGCGCGCTACGCGAGCGGTGCGTCCGCGCC 200
QY 184 CAGGTGCTGGCTTGACCTCGACGAGCGCTTGGCGGCGGAGTCTGGCCCGCGCTC 243
Db 201 CCGCTGTACTCTCGCGCGCTCTCGAGTACCTCGCGCGCGAGTCTGGAGTCCCGCGC 260
QY 244 ACCGCGCCAGGACCCCAACCATGACCTGACGAGGAGAAATCGAGTGTCTCGCACCGAC 303
Db 261 AAGCGCGCAAGAGCAACAGAGAGCGCATGTCGCGCGCACCTGTCTCTCGCATC 320
QY 304 CCGAAGCTGGAGAGCATGAGGCGCGCATATCGCCCTAATCTGTGAGGCGCGCTAG 363
Db 321 CGCAACGACGAGAGTCTCGGCAAGTGTCTCGCGCGCTCACCATCGCGACGCGCGCTG 380
QY 364 CGCGACAGCGCGCGCGCATCGAGGAACCAAGCGCTCATCGA 407
Db 381 CTGCCCCAATCAACCCCGTGTCTCCCAAGAGGCGCTCGA 424

RESULT 12

CK212590/c
LOCUS CK212590.1
DEFINITION FGAS024469 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum
aestivum cDNA, mRNA sequence.
ACCESSION CK212590
VERSION CK212590.1 GI:39618694
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 1035)
REFERENCE Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Murak, I., Nilsson, D.,
Penniket, C., Roach, J.B., and Sarhan, F., 2003.
Functional Genomics of Abiotic Stress in Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769

Fax: 306 966 2033
Email: fgas.ests@cs.usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [21,782].
Plate: L6B006 row: E column: 04.
Location/Qualifiers

FEATURES

1..1035
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown
(50%) and leaf (50%) tissues from wheat cultivar Norstar
after short exposure times to low temperature in the light
and in the dark. 12 mRNA populations were combined before
constructing the library. The first 6 populations: After 7
days of growth at 20C from wheat cultivar Norstar after
short exposure times to low temperature in the light and
in the dark. 12 mRNA populations were combined before
constructing the library. The first 6 populations: After 7
days of growth at 20, wheat plants were transferred to 4C
in the light. 1cm crown sections and green leaf tissue were
separately harvested after 1, 3, and 6 hours of low
temperature exposure. The last 6 populations: After 7 days
of growth at 20C, wheat plants were transferred to 4C in
the dark. 1cm crown sections and green leaf tissue were
separately harvested after 1, 3, and 6 hours of low
temperature exposure. First strand synthesis in this
library was done in the presence of methylated dCTP
thereby protecting from internal cleavage with NotI. In
addition, this library used a primer for second strand
synthesis that annealed to an artificial sequence (RNA
oligo) added before first strand synthesis. Therefore when
sequences from EST generated from this library will be
masked for vector and adaptor sequences, an additional
masking step will have to be included to mask this RNA
oligo that is common to all clones (sequence
CGACTGGACGAGGACACTGACATGGACTGAGGAGTAGAAA)."

ORIGIN

Query Match 14.6%; Score 62.4; DB 14; Length 1035;
Best Local Similarity 50.9%; Pred. No. 19;
Matches 175; Conservative 0; Mismatches 166; Indels 3; Gaps 1;
QY 67 GCGGACTGACCAAGAGCGAGTTGCCAGGCGCATCCAGAGG---ACGGGCCACCGTC 123
Db 718 GCGCGAGAGAGGAAGAGCGGTGACCGGTCCTCAAGGCGCGCTCCAGTTCCTCCGTC 659
QY 124 GCGCGTGGGAGGACGCAAGAACCGCGCGACGCGGACCTCGTTGCCCGCTGCC 183
Db 658 GCGCGCATCGGGCGCTTCTCAAGAGGCGGTCTACGCGAGCGCTCGGCTCCGCGCC 599
QY 184 CAGGTGCTCGGCTCGACCTCGACGAAGCCCTCGCGCGCGAGGTCTGGCGCCCGCGTC 243
Db 598 CCGCTGTACTCTCGCGCGCTCTCGAGTACCTCGCGCGCGAGGTCTGGAGCTCGCGCGC 539
QY 244 ACCGCGCAGGACCCCAACCATGACCTGACGAGGAATCGAGCTGGTCCGACCGAC 303
Db 538 AACCGCGCAAGGACCAAGAGACCGCATCTGCGCGCCACCTGCTCTCGCATC 479
QY 304 CCGAAGCTGGACGAGGACATGAAGCGCGCATCATCGCCCTAATCTGTGAGCGCGTGA 363
Db 478 CGCAACGACGAGGAGCTCGGCAAGTGTCTCGCGCGCTCACCATCGCGACGCGCGCTG 419
QY 364 CGCGACAGGCGCGCGCGATCGAGGAACCAAGCGGCTCATCGA 407
Db 418 CTCCCCAATCAACCCCGTGTCTCTCCCCAAGAGGCGCTCGA 375

RESULT 13

Db 124 CCTCGCGACCTCGTCACCGTCTTCCCTGCGCCCGCGCTCAAGGACGCCGCCCGCTGCGC 183
 Qy 264 CATGACCTGACGAGGAAATCGAGCTGGTCGCGACCGACCCCAAGCTGGAGGAGCAT 323
 Db 184 CATCTCCCTTCGAGGACTCGTCGACGGGATCAGCGGCAACCTCTTCGACGCTAGGT 243
 Qy 324 GAAGCGGCGCATCATCGCCCTAATCCTGAGGCGCGTGGAGCGGCAAGCGGCGCGCAT 383
 Db 244 CAAGCCCTACTTCAATTGCGCGCTGCGCGCGTGGCGAAGCGCGCACCGGTTCTGTGTGA 303
 Qy 384 CG 385
 Db 304 CG 305

RESULT 15
 BE400165 424 bp mRNA linear EST 21-JUL-2000
 LOCUS AWB011.D10F000328 ITEC AWB Wheat Meiotic Stage Library Triticum
 DEFINITION aestivum cDNA clone AWB011.D10, mRNA sequence.

ACCESSION BE400165

VERSION BE400165.1 GI:9359633

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.

AUTHORS 1 (bases 1 to 424)

Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,
 Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J.,
 Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P.,
 Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G.,
 Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.
 International Triticeae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticeae

Unpublished (2000)

JOURNAL

COMMENT Contact: Langridge P
 Special Research Centre, Basic and Applied Plant Molecular Biology
 Waite Campus, University of Adelaide, Glen Osmond SA 5064 AUSTRALIA
 Tel: 61 8 8303 7368

Fax: 61 8 8303 7102

Email: plangridge@waite.adelaide.edu.au

International Triticeae EST Cooperative (ITEC)

http://wheat.pw.usda.gov/genome.

Location/Qualifiers

FEATURES

source

1..424 /organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="AWB011.D10"

/tissue_type="excised florets"

/dev_stage="meiotic stage no later than metaphase I"

/clone_lib="ITEC AWB Wheat Meiotic Stage Library"

/notes="Vector: pSport 1 (Life Technologies cat. no.

18248-013); Site 1: SalI; Library constructed in pSport 1.

Directionally cloned using the Superscript Plasmid System

for cDNA synthesis and plasmid cloning. M13 Reverse

sequencing primer used to obtain 5' sequence data. 1.4

Kbp average insert size."

ORIGIN

Query Match 14.5%; Score 61.6; DB 10; Length 424;

Best Local Similarity 52.3%; Pred.No. 23; Indels 3; Gaps 1;

Matches 161; Conservative 0; Mismatches 144; Indels 3; Gaps 1;

Qy 67 GCGGACTGACGAGCGAGTTGGCCAGCGCGCATCCAGAAG---ACGGGCCACCGTC 123

Db 91 GCGGAGAGGAGGAGGCGGTGACCCCGTCCACCAAGCGCGCTCCAGTTCGCCGTC 150

Qy 124 GCGCGTGGAGAGCGGCAAGACCGCGCGGACGCGGACCTGCTTCCCGCGCTCGCC 183
 Db 151 GCGCGATCGGGCGCTTCTCAAGAGGGCGGTACGCGAGCGCGTTCGCTCCGCGCC 210
 Qy 184 CAGGTGCTCGGCGCTCGACCTCGACGAGCCCTCGCCCGCGCAGGTCTGGCCCGCGCGTC 243
 Db 211 CCGGTCTACCTCGCGCGCTCTCGAGTACCTCGCGCGCGAGGTGCTGGAGTTCGCGCGC 270
 Qy 244 ACCCGCGCAGGACCCCAACCATGGACCTGGACGAGGAAATCGAGTGGTCCGACCGAC 303
 Db 271 AACGCGCCCAAGACAACAAGAAGACCCCGCATCGTGGCGCGCCACCTGCTCTCGCCATC 330
 Qy 304 CCCAAGCTGGACGAGGACATGAAGCGCGCGCATCATGCCCCCTAATCTTGGAGCGCCGTGAG 363
 Db 331 CGCAACGACGAGGAGCTCGGCAAGCTGCTCGCCGGGTACCATCGCGCACGCGGCGGTG 390
 Qy 364 CGGACAA 371
 Db 391 CTCGCCAA 398

Search completed: August 6, 2004, 13:39:02

Job time : 1653.94 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM nucleic - nucleic search, using sw model
Run on: August 6, 2004, 08:50:04 ; Search time 156.122 Seconds
(without alignments)
9439.175 Million cell updates/sec

Title: US-09-855-340A-3
Perfect score: 34
Sequence: 1 ccccgtaagggttcattccatccagtcacccg 34

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	34	100.0	34	6	AX338972	AX338972 Sequence
2	34	100.0	145	1	AY150033	AY150033 Micromono
3	34	100.0	247	6	AX338975	AX338975 Sequence
4	34	100.0	2025	7	AY150027	AY150027 Bacteriop
5	33	97.1	143	1	AY150032	AY150032 Micromono
6	33	97.1	260	6	AX338978	AX338978 Sequence
7	25	73.5	143	1	AY150031	AY150031 Micromono
8	25	73.5	145	1	AY150028	AY150028 Micromono
9	25	73.5	145	1	AY150029	AY150029 Micromono
10	25	73.5	241	6	AX338973	AX338973 Sequence
11	25	73.5	243	6	AX338974	AX338974 Sequence
12	25	73.5	315	6	AX338977	AX338977 Sequence
c 13	24.6	72.4	34806	1	AF481102	AF481102 Candidatu
14	24	70.6	143	1	AY150030	AY150030 Micromono
15	24	70.6	255	6	AX338976	AX338976 Sequence
16	23.6	69.4	15708	3	AF051097	AF051097 Balanoglo
17	22.4	65.9	180	1	MYCTGH	LY5239 Mycoplasma
18	22.4	65.9	10252	1	U39716	U39716 Mycoplasma
19	22.4	65.9	10640	1	AE000029	AE000029 Mycoplasma
20	22.4	65.9	110000	6	AR300198	Continuation (5 of
c 21	22.4	65.9	155932	1	CJ11168X5	AL139079 Campyloba
22	22.4	65.9	224279	10	AC102135	AC102135 Mus muscu
c 23	22.2	65.3	215065	2	AC103291	AC103291 Rattus no
c 24	22.2	65.3	222507	2	AC112440	AC112440 Rattus no
c 25	22.2	65.3	226214	2	AC126653	AC126653 Rattus no
c 26	22	64.7	236921	2	AC094151	AC094151 Rattus no
c 27	22	64.7	241131	2	AC118438	AC118438 Rattus no
28	21.8	64.1	620	8	AF369748	AF369748 Populus d
29	21.8	64.1	174470	10	AC114820	AC114820 Mus muscu
30	21.8	64.1	205196	2	AC131173	AC131173 Rattus no
31	21.8	64.1	222714	2	AC116514	AC116514 Mus muscu
32	21.8	64.1	261157	2	AC103330	AC103330 Rattus no
33	21.8	64.1	284749	2	AC113251	AC113251 Rattus no
34	21.4	62.9	236458	2	AC107428	AC107428 Rattus no
35	21.4	62.9	271197	2	AC117974	AC117974 Rattus no
36	21.2	62.4	110970	10	AC004807	AC004807 Mus muscu
37	21.2	62.4	150986	2	AC023802	AC023802 Mus muscu
38	21.2	62.4	151437	2	AC116850	AC116850 Mus muscu
c 39	21.2	62.4	181721	2	AC102230	AC102230 Mus muscu
c 40	21.2	62.4	194695	2	AC026375	AC026375 Mus muscu
c 41	21.2	62.4	197338	2	AC108265	AC108265 Rattus no
c 42	21.2	62.4	233242	10	AL591070	AL591070 Mouse DNA
c 43	21.2	62.4	247759	2	AC106669	AC106669 Rattus no
c 44	21	61.8	110000	2	AC110649_0	AC110649 Rattus no
c 45	21	61.8	146984	2	AC144875	AC144875 Pongo pyg

ALIGNMENTS

RESULT 1
AX338972
LOCUS AX338972 34 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3 from Patent WO0187936.
ACCESSION AX338972
VERSION AX338972.1 GI:18129108
KEYWORDS Micromonospora carbonacea
SOURCE Micromonospora carbonacea
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1 Hosted, T.J. and Horan, A.C.
AUTHORS Isolation of Micromonospora carbonacea var africana pmlp1 integrase
TITLE and use of integrating function for site-specific integration into

```

Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 3 22-NOV-2001;
SCHERING CORPORATION (US)
FEATURES
    source
        1. .34
            /organism="Micromonospora carbonacea"
            /mol_type="unassigned DNA"
            /db_xref="taxon:47853"
ORIGIN
Query Match      100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 34
    |||||
Db 1 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 34
    |||||
RESULT 2
AV150033      145 bp DNA linear BCT 02-SEP-2003
LOCUS
DEFINITION
Micromonospora sp. ATCC 39149 attR region.
ACCESSION
AV150033
VERSION
AV150033.1 GI:28630429
KEYWORDS
SOURCE
Micromonospora sp. ATCC 39149
ORGANISM
Micromonospora sp. ATCC 39149
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1 (Bases 1 to 145)
Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE
Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL
Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED
12949170
REFERENCE
2 (Bases 1 to 145)
Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
AUTHORS
Direct Submission
TITLE
Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
JOURNAL
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        56. .145
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="ATCC:39149"
            /db_xref="taxon:219305"
            /focus
            /notes="deposited in ATCC as Micromonospora carbonacea var.
africana Waitz et al."
        1. .55
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        1. .145
            /notes="attR region"
        55. .56
            /notes="site of pSPRH840 recombination"
        56. .79
            /notes="pSPRH840 attachment site; attB/attP identity
element"
    stem_loop
        join(85. .98,103. .116)
ORIGIN
Query Match      100.0%; Score 34; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 34
    |||||

```

```

Db 47 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 80
RESULT 3
AX338975      247 bp DNA linear PAT 09-JAN-2002
LOCUS
DEFINITION
Sequence 6 from Patent WO0187936.
ACCESSION
AX338975
VERSION
AX338975.1 GI:18129111
KEYWORDS
Micromonospora carbonacea
SOURCE
Micromonospora carbonacea
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE
1
Hosted,T.J. and Horan,A.C.
TITLE
Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into
Micromonospora halophitica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 6 22-NOV-2001;
SCHERING CORPORATION (US)
JOURNAL
Location/Qualifiers
FEATURES
    source
        1. .247
            /organism="Micromonospora carbonacea"
            /mol_type="unassigned DNA"
            /db_xref="taxon:47853"
ORIGIN
Query Match      100.0%; Score 34; DB 6; Length 247;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 34
    |||||
Db 92 CCGCGTACGGGTTCAATTCCTCCATCAGTCACCG 125
    |||||
RESULT 4
AX150027      2025 bp DNA linear PHG 02-SEP-2003
LOCUS
DEFINITION
Bacteriophage pMLP1 att/int region.
ACCESSION
AX150027
VERSION
AX150027.1 GI:28630421
KEYWORDS
Bacteriophage pMLP1
SOURCE
Bacteriophage pMLP1
ORGANISM
Viruses.
REFERENCE
1 (Bases 1 to 2025)
Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE
Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL
Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED
12949170
REFERENCE
2 (Bases 1 to 2025)
Hosted,T.J. Jr., Alexander,D.C. and Hewitt,D.D.
AUTHORS
Direct Submission
TITLE
Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
JOURNAL
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1. .2025
            /organism="Bacteriophage pMLP1"
            /mol_type="genomic DNA"
            /specific_host="Micromonospora sp. ATCC 39149"
            /db_xref="taxon:219292"
            /note="host is deposited in ATCC as Micromonospora
carbonacea var. africana Waitz et al."
        1. .2025
            /note="att/int region"
        197. .622
            /note="xis"
        /codon_start=1
    misc_feature
    CDS

```

```
/transl_table=11
/product="excisionase"
/protein_id="AA046044.1"
/db_xref="GI:28630422"
/translation="WRNTPGICRGRTWAAYVLTABERAGLTGKSELARRIQKDRATVGRW
EDGNRPDADLVARVAQVGLDDDEALAAAGLRPGVTPPATFINDDEIELELVRTDP
KLDEDMKRIITALILERRERDKAAAEETKELIDLFRRS"
/gene="1806"
/size="628"
/feature="intm"
/feature="intm"
/feature="intm"
/codon_start=1
/transl_table=11
/product="integrase"
/protein_id="AA046045.1"
/db_xref="GI:28630423"
/translation="WMIKNGEVVHRLVVRGKVTIQTGYPTKTSKAMVOPRAEQ
LQGNALMPRGQITLADVGNWPSYKTLKPTAVNSEGNIRNHLPLILGHLTLDL
DQVQVQWNLDEAGVWPSTGRGRPLAAKTLISNCHGLLHICGNAIAAKIRLN
PCSTMPLRPREPKVKFLUSDPEIGLITLALPPHWRPLVNLVATGLRWGEAIGURGR
VLLAARPLRTVQLQELASTGELVFQSPKTKAGRTVSTFTVALLLTPLIAGKRS
DEVPTAPKGMVTRNFRRIWVKAEEAGLPLRIHDLRTHAAILISAGRPLSATIS
RRLGHSSIAVTDLLYCHLREEVDEGILAAIEEAVAGVRAEDLEAELELDELTDLVLAADAA
"
stem_loop
misc_feature
stem_loop
ORIGIN
Query Match 100.0%; Score 34; DB 7; Length 2025;
Best Local Similarity 100.0%; Pred. No. 0.00016; Indels 0; Gaps 0;
Matches 34; Conservative 0; Mismatches 0;

Qy 1 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 34
|||||
Db 1916 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 1949

RESULT 5
AY150032 AY150032 143 bp DNA linear BCT 02-SEP-2003
LOCUS Micromonospora nigra attR region.
DEFINITION Micromonospora nigra attR region.
ACCESSION AY150032
VERSION AY150032.1 GI:28630428
KEYWORDS Micromonospora nigra
SOURCE Micromonospora nigra
ORGANISM Micromonospora nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1 (bases 1 to 143)
AUTHORS Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED 12949170
REFERENCE 2 (bases 1 to 143)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
Location/Qualifiers
source
56..143
/organism="Micromonospora nigra"
/mol_type="genomic DNA"
/strain="ATCC 33088"
/db_xref="ATCC:33088"
/db_xref="taxon:145857"
/focus
1..55
/organism="Integration vector pSPRH840"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:219291"
1..143
/notes="attR region"
55..56
/notes="site of pSPRH840 recombination"
56..79
/notes="pSPRH840 attachment site; attB/attP identity
element"
join(94..105,110..121)
ORIGIN
Query Match 97.1%; Score 33; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 33
|||||
Db 47 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 79

RESULT 6
AX338978 AX338978 260 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 9 from Patent WO0187936.
DEFINITION AX338978
ACCESSION AX338978
VERSION AX338978.1 GI:18129114
KEYWORDS Micromonospora halophytica
SOURCE Micromonospora halophytica
ORGANISM Micromonospora halophytica
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1
AUTHORS Hosted,T.J. and Horan,A.C.
TITLE Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into
Micromonospora halophytica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 9 22-NOV-2001;
SCHERING CORPORATION (US)
JOURNAL
FEATURES
Location/Qualifiers
source
1..260
/organism="Micromonospora halophytica"
/mol_type="unassigned DNA"
/db_xref="taxon:47864"
ORIGIN
Query Match 97.1%; Score 33; DB 6; Length 260;
Best Local Similarity 100.0%; Pred. No. 0.00046;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 33
|||||
Db 92 CCCCGTACGGGTTCAATTCATCCATCAGTCACCC 124

RESULT 7
AY150031 AY150031 143 bp DNA linear BCT 02-SEP-2003
LOCUS Micromonospora nigra attL region.
DEFINITION Micromonospora nigra attL region.
ACCESSION AY150031
VERSION AY150031.1 GI:28630427
KEYWORDS Micromonospora nigra
SOURCE Micromonospora nigra
ORGANISM Micromonospora nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE 1 (bases 1 to 143)
AUTHORS Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
```

```

PUBMED 12949170
REFERENCE 2 (bases 1 to 143)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..79
            /organism="Micromonospora nigra"
            /mol_type="genomic DNA"
            /strain="ATCC 33088"
            /db_xref="taxon:33088"
            /db_xref="taxon:145857"
            /focus
        80..143
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..143
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..79
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 79..80
            /note="site of pSPRH840 recombination"
        stem_loop join(100..118,122..138)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

PUBMED 12949170
REFERENCE 2 (bases 1 to 143)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..79
            /organism="Micromonospora nigra"
            /mol_type="genomic DNA"
            /strain="ATCC 33088"
            /db_xref="taxon:33088"
            /db_xref="taxon:145857"
            /focus
        80..143
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..143
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..79
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 79..80
            /note="site of pSPRH840 recombination"
        stem_loop join(100..118,122..138)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

PUBMED 12949170
REFERENCE 2 (bases 1 to 143)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..80
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="taxon:39149"
            /db_xref="taxon:219305"
            /focus
            /note="deposited in ATCC as Micromonospora carbonacea var.
            africana Waitz et al."
        81..145
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..145
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..80
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 80..81
            /note="site of pSPRH840 recombination"
        stem_loop join(100..117,122..141)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

```

```

PUBMED 12949170
REFERENCE 2 (bases 1 to 145)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..80
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="taxon:39149"
            /db_xref="taxon:219305"
            /focus
            /note="deposited in ATCC as Micromonospora carbonacea var.
            africana Waitz et al."
        81..145
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..145
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..80
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 80..81
            /note="site of pSPRH840 recombination"
        stem_loop join(100..117,122..141)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

PUBMED 12949170
REFERENCE 2 (bases 1 to 145)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..80
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="taxon:39149"
            /db_xref="taxon:219305"
            /focus
            /note="deposited in ATCC as Micromonospora carbonacea var.
            africana Waitz et al."
        81..145
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..145
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..80
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 80..81
            /note="site of pSPRH840 recombination"
        stem_loop join(100..117,122..141)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

PUBMED 12949170
REFERENCE 2 (bases 1 to 145)
AUTHORS Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
    source
        1..80
            /organism="Micromonospora sp. ATCC 39149"
            /mol_type="genomic DNA"
            /strain="ATCC 39149"
            /db_xref="taxon:39149"
            /db_xref="taxon:219305"
            /focus
            /note="deposited in ATCC as Micromonospora carbonacea var.
            africana Waitz et al."
        81..145
            /organism="Integration vector pSPRH840"
            /mol_type="genomic DNA"
            /db_xref="taxon:219291"
        misc_feature 1..145
        tRNA 5..79
            /product="tRNA-His"
        misc_feature 56..80
            /note="pSPRH840 attachment site; attB/attP identity
            element"
        misc_recomb 80..81
            /note="site of pSPRH840 recombination"
        stem_loop join(100..117,122..141)
ORIGIN
Query Match 73.5%; Score 25; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAAATCCCATCAGTCACCCG 34
Db 56 GGGTTCAAATCCCATCAGTCACCCG 80

```

Db 56 GGGTTCAATTCCTCAGTCACCG 80

AX338973 241 bp DNA linear PAT 09-JAN-2002

LOCUS AX338973

DEFINITION Sequence 4 from Patent WO0187936.

ACCESSION AX338973

VERSION AX338973.1 GI:18129109

KEYWORDS

ORGANISM Micromonospora carbonacea

SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 4 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .241

/organism="Micromonospora carbonacea"

/mol_type="unassigned DNA"

/db_xref="taxon:47853"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 241;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 95 GGGTTCAATTCCTCAGTCACCG 119

RESULT 11

AX338974

LOCUS AX338974

DEFINITION Sequence 5 from Patent WO0187936.

ACCESSION AX338974

VERSION AX338974.1 GI:18129110

KEYWORDS

ORGANISM Micromonospora carbonacea

SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 5 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .243

/organism="Micromonospora carbonacea"

/mol_type="unassigned DNA"

/db_xref="taxon:47853"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 243;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 113 GGGTTCAATTCCTCAGTCACCG 137

RESULT 12

AX338977

LOCUS AX338977

DEFINITION Sequence 8 from Patent WO0187936.

ACCESSION AX338977

VERSION AX338977.1 GI:18129113

KEYWORDS

ORGANISM Micromonospora halophytica

SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micromonosporaceae; Micromonospora.

REFERENCE 1

AUTHORS Hosted,T.J. and Horan,A.C.

TITLE Isolation of Micromonospora carbonacea var africana pmpl1 integrase and use of integrating function for site-specific integration into Micromonospora halophytica and Micromonospora carbonacea chromosome Patent: WO 0187936-A 8 22-NOV-2001;

JOURNAL SCHERING CORPORATION (US)

FEATURES

source Location/Qualifiers

1. .315

/organism="Micromonospora halophytica"

/mol_type="unassigned DNA"

/db_xref="taxon:47864"

ORIGIN

Query Match 73.5%; Score 25; DB 6; Length 315;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCAGTCACCG 34

Db 96 GGGTTCAATTCCTCAGTCACCG 120

RESULT 13

AF481102/c

LOCUS AF481102

DEFINITION Candidatus Tremblaya princeps mvin DNA linear BCT 07-JUL-2002

ribosomal protein L13 (rpl13), ribosomal protein S9 (rps9), putative protein HesB (hesB), dihydroxyacid dehydrase (ilvd), 5-enolpyruvylshikimate-3-phosphate synthase (aroA), ribosomal protein S1 (rps1), ribosomal protein S16 (rps16), ribosomal protein L19 (rpl19), 5, 10-methylenetetrahydrofolate reductase (metF), chaperone Hsp60 (groEL), chaperone Hsp10 (groES), DNA polymerase alpha subunit (dnaE), valine sensitive acetolactate synthase III subunit (ilvi), ketol-acid reductoisomerase (ilvC), alpha-isopropylmalate synthase (leuA), and ribosomal protein S15 (rps15) genes, complete cds; 16S ribosomal RNA, 23S ribosomal RNA, and 5S ribosomal RNA genes, complete sequence; putative protein Yabc (yabc), DNA polymerase III epsilon subunit (dnaQ), cell division protein (ftsJ), arginosuccinate lyase (argH), and ribosomal protein L31 (rpl31) genes, complete cds; and unknown genes.

ACCESSION AF481102

VERSION AF481102.1 GI:21702567

KEYWORDS

ORGANISM Candidatus Tremblaya princeps

SOURCE Candidatus Tremblaya Princeps

REFERENCE 1

AUTHORS Baumann,L., Thao,M.L., Hess,J.M., Johnson,M.W. and Baumann,P.

TITLE The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects

JOURNAL Appl. Environ. Microbiol. 68 (7), 3198-3205 (2002)

MEDLINE 22083449

PUBMED 12089995

REFERENCE 2

(bases 1 to 34806)

AUTHORS Baumann,L., Thao,M.L., Hess,J.M. and Baumann,P.

TITLE Direct Submission

JOURNAL Submitted (06-FEB-2002) Microbiology, University of California, One Shields Ave., Davis, CA 95616-8665, USA

FEATURES

source Location/Qualifiers

1. .34806

/organism="Candidatus Tremblaya princeps"

/mol_type="genomic DNA"
/specific_host="Dysmicoccus brevipes"
/db_xref="taxon:189385"
/note="endosymbiont"
<1..843
/gene="mvin"
<1..843
/gene="mvin"
/note="putative virulence protein"
/codon_start=1
/transl_table=11
/product="mvin"
/protein_id="AA075965.1"
/db_xref="GI:21702568"
/translation="EFSTGLYSGSLSLPSSRSWASGDCAGYAVLMDGLSIVAL
AGAPNALLFTLLPELTSASFNYSGRFDHTVQWVAPLAAVPGMLGMLLRTLSQGL
YARNVSTVPLASLLICLSQAAGRLMLPMGHAGLMSSTIGAMNLAALLRSLSQ
GAYATPRWASLCARISAAVACTILSAIARRVDWGMVSURPELRAALASVIAAF
AYAFILKGLGLMOMWREASCAASNTVPSNQYKSRAGQARALAVAGNRKLGP
FRQA"
1136..1510
/gene="rpl13"
1136..1510
/gene="rpl13"
/codon_start=1
/transl_table=11
/product="ribosomal protein L13"
/protein_id="AA075966.1"
/db_xref="GI:21702569"
/translation="MKDNGQHVLDARGMVLGRVAVTLAGLLRGKYKASYAHNAAPHNA
VTVLNCHEVRISGKAVGVYVYHSGHPCGLRRTKLHDLMLRSPATVLRRAVHGMLPR
NLDRPMSMLVHANEGRAR"
1507..1893
/gene="rps9"
1507..1893
/gene="rps9"
/codon_start=1
/transl_table=11
/product="ribosomal protein S9"
/protein_id="AA075967.1"
/db_xref="GI:21702570"
/translation="MTGWRYGRRKRSVARVFLKAAAGTAVVNGMPLGLYFRSAH
AGAAAPLHVTPVAVKANVQGGIGGAGAVRLGISALACLSPLRLTIASARLL
TDQRVERKKQLKARRRQFSKR"
complement (1899..2225)
/gene="hesB"
1899..2225
/codon_start=1
/transl_table=11
/product="putative protein HesB"
/protein_id="AA075968.1"
/db_xref="GI:21702571"
/translation="MLQINCAVLELRGMVACHGKSCVPLRVFVKGSGCKGMRHFCFL
ASRTVGDAKMTLGRVLVRLDPLSLMYLRGSIYDCTGVGDARLVVYQPSAGWCSCG
ASPTPV"
2317..3981
/gene="ilvD"
2317..3981
/gene="ilvD"
/codon_start=1
/transl_table=11
/product="dihydroxyacid dehydrase"
/protein_id="AA075969.1"
/db_xref="GI:21702572"
/translation="MNTNGRRITHGFRSPNRMILCAMGYDQADFAPKPIIGVSGYST
ITPCAGIOVVDAAKAMLRHGMATQVFGIPTISDGLSMGTGPMGRYSLVREAVADC
IFACAGQGLLVVGGCDKXNLPGCMIGLLRANIPIGLYGTTLPGYVWGAELTVV
SFEACVAGRCQMSIDANREVERHACPTGCGMYTANTMSAFEGALGWSLYSST
APSGEGELISHSARSILINAVRGIRPRDVYTHRSIRKNAVMVAVGGSTNAVLYH
LAIAAARGALSISLHVELIRRVFVPCNNKPSGLSHLADLHAGGVPVRVWHELALGL
IDESCLITGRTIGALLAARHKGSTVTLTNVLYRTGLVLSGNMNSGNVAK
TSGLSILLSGTAVFRSEACVQALINGCRVIGDVVLIYLGPKGPGFEMPLGFTA
ALVGMGLGQSACLITDGRFSGGTWGLVVGHSPEAAVGGSIALVRENGDFITVDLRNNS

LHLHIDAWSLAARRAARWLPTCLVTEGLLRKYKTHDGVGQSHNGAIAA"
4467..5498
/note="ORF-A"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA075970.1"
/db_xref="GI:21702573"
/translation="MOSLATMNLGALFPVLTLCSSAWHPIPCPVSNPCPRNKRRT
AIAVALSFAPNLLARMSRHCNFRFVLYHEQSMGVNDKTHHILIFLAACLREMSG
HPAHRATGKALACACLLQGMHEIROVQLESQPVARCGRGKSNASTACIACVADGAG
GDMFATLISRDLPAYAKQDGPPIPRKLEAGELMVAECVPMISNLRVDRVIRTEOLI
PIVESNITANRNLTPSORARVLMWRCVDSVHKAAAYAPFAGRGFWNAVIHELWAY
SRYTTELDITDEFYSSYTAGSKYLKRRQVGRDVRDLVKRCTLQHLASLYRIRAE
AIEMLGLIIG"
5499..6500
/note="ORF-B"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA075971.1"
/db_xref="GI:21702574"
/translation="MPSFHTELRYINGHDGLSACAARLGVASRIGDMKRAAGMCS
YDARTRIETVAREVVVVDVYGGRIAPAPLSLWRCIMNGARSVEPECLALHLPMTCH
SERATMALRQSORQVQFASFGACSVACRAGCTILPVHNTLSGAVSAAIRVCSVLRL
LAALCSVAVMHPAPSSGRVAVTAQVAKHALWCSSWFCDSRMVASVARSASEPSI
AAAQACACATTPALARAMRGAALAHSENNRTSFVACSCAGMRSVIGSRVAVMS
LDSRCCEVHCALLLIALRHSPSCHVLVIEARAGGAFFVARSRLRVLPQAGCPCQ"
6485..7783
/gene="aroA"
6485..7783
/gene="aroA"
/codon_start=1
/transl_table=11
/product="5-enolpyruvylshikimate-3-phosphate synthase"
/protein_id="AA075972.1"
/db_xref="GI:21702575"
/translation="MPLVRSARVIRRRHRRGLALVPTSKSASNRREVACAFVSLKATQ
LSGRTLHSDDTAVLNAHSAFGILAGHGRTLELGPCKHGLLVAELVYVSGTTPAR
LVGASACRRPRHALVHGDRPMRRPALATVGLASLAKVLFHFAVQSLPVAIGTTCN
LKNEAPRLPMSSSQFITSASVASHGIVHKDTPACGOLTSRLYVASTIDIMGLFG
MCSEGGKQRYAHAPASMDCEYFGSTDFDFTWSASYPTALCCSGMGMWRLCGVDHASQ
TADAMASMGCAWQCQCDGNAMRMAAHKHAFTPRLLCCSAPDGMATVATISLALAG
ASDAVGLSYRYKETDRVLAVALELRMGVAADCGMGSMACLSLUTWHGRADVSTIR
DHRIAMCCSLMHAAGCOVRTSPQCVCCTPYDYLHECAGT"
7899..9455
/gene="rps1"
7899..9455
/gene="rps1"
/codon_start=1
/transl_table=11
/product="ribosomal protein S1"
/protein_id="AA075973.1"
/db_xref="GI:21702576"
/translation="MHDHIFELREYCKYDAGPGSIVAAARVVAISSNHVVLRLKLS
SRVPLGEFVSPQDRLEVKVGDVTLAYVERDAYDEGTVLRLSKSRARVAVWGLDSAH
RTRGTINGTVGRVKGMTAVVNGMEAFIFGSLIDTRPKDAERYGRRLRFRHVLKLD
RKRGVLSRRSVIERRQANRNRLILASGDGVQGVKNITDYGAFIDLGLGLL
HITDISWRIRHPSQVLSVGDRTIVKVLKFDTRKTVSLGKLGKQLHPPWAGIIGRPVA
GSRVGRVTNIADYQVLFVEIERGIEGLHVSEMKSGKTSILRHFRVQYVNVTVLD
VSKTRRIISLGMKHGTLNPLWTFYKGRKNDRMGAIIRSPDVGLVDLPMPGVHGVVC
MSXDLRPLRLALCSGVVEARVVGVDVNRGRISLASGALGPLGVRIILAPGPIRPGA
FVRMAPNRECLINVRAPCAAAQAQALRGAVHARLLLSAISIRSPRSVACAIRGMAR
ANHGLVPLCP"
9278..9724
/gene="rps16"
9278..9724
/gene="rps16"
/codon_start=1
/transl_table=11
/product="ribosomal protein S16"
/protein_id="AA075974.1"
/db_xref="GI:21702577"

```

/translation="MLPORASSAVRCQCGPVRSSSRSCDAAKCYFTYSSLCMRHS
RYIASKPWACSLAIRRCGARGRPAYQVVVADSRKRGVFLCRVGYINPRLUKSAH
IDTVMLRLWTGRTGAPTTVAELLHRHAAGCTAAARNSGVCPHMQ"
9706..110020
/gene="rpl19"
9706..110020
/gene="rpl19"
/codon_start=1
/transl_table=11
/product="ribosomal protein L19"
/protein_id="RAM75975.1"
/db_xref="GI:21702578"
/translation="MSTAIIRVLEARTGMAATDAPLPYPRIGDVAISGTTGRGKGY
APEGVVSRRSALNTSIVVREHAGFCVDLALRLHSHDVPARSARSTSCAAPACVC
RP"

Query Match      72.4%; Score 24.6; DB 1; Length 34806;
Best Local Similarity 87.1%; Pred. No. 3.6;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CCGGTACGGTTCAATCCCATCAGTCACCC 33
Db 32511 CAGGTATGGTTCGAGTCCCATCAGTCACCC 32481

RESULT 14
AY150030      143 bp      DNA      linear      BCT 02-SEP-2003
DEFINITION   Micromonospora nigra attB region.
ACCESSION   AY150030
VERSION      AY150030.1 GI:28630426
KEYWORDS
SOURCE       Micromonospora nigra
ORGANISM     Micromonospora nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE    1 (bases 1 to 143)
AUTHORS      Alexander,D.C., Devlin,D.J., Hewitt,D.D., Horan,A.C. and
Hosted,T.J.
TITLE        Development of the Micromonospora carbonacea var. africana ATCC
39149 bacteriophage pMLP1 integrase for site-specific integration
in Micromonospora spp
JOURNAL      Microbiology (Reading, Engl.) 149 (Pt 9), 2443-2453 (2003)
PUBMED      12949170
REFERENCE    2 (bases 1 to 143)
AUTHORS      Hosted,T.J. Jr., Hewitt,D.D. and Alexander,D.C.
TITLE        Direct Submission
JOURNAL      Submitted (12-SEP-2002) New Lead Discovery, Schering-Plough
Research, 2015 Galloping Hill Road, Kenilworth, NJ 07033, USA
FEATURES
source      1..143
            /organism="Micromonospora nigra"
            /mol_type="genomic DNA"
            /strain="ATCC 33088"
            /db_xref="ATCC:33088"
            /db_xref="taxon:145857"
misc_feature 1..143
            /note="attB region"
tRNA         5..79
            /product="tRNA-His"
misc_feature 56..80
            /note="attachment site; attB/attP identity element"
stem_loop   join(94..105,110..121)
ORIGIN
Query Match      70.6%; Score 24; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCATCAGTCACCC 33
Db 56 GGGTTCAATTCCTCATCAGTCACCC 79

/translation="WLPORASSAVRCQCGPVRSSSRSCDAAKCYFTYSSLCMRHS
RYIASKPWACSLAIRRCGARGRPAYQVVVADSRKRGVFLCRVGYINPRLUKSAH
IDTVMLRLWTGRTGAPTTVAELLHRHAAGCTAAARNSGVCPHMQ"
9706..110020
/gene="rpl19"
9706..110020
/gene="rpl19"
/codon_start=1
/transl_table=11
/product="ribosomal protein L19"
/protein_id="RAM75975.1"
/db_xref="GI:21702578"
/translation="MSTAIIRVLEARTGMAATDAPLPYPRIGDVAISGTTGRGKGY
APEGVVSRRSALNTSIVVREHAGFCVDLALRLHSHDVPARSARSTSCAAPACVC
RP"

Query Match      72.4%; Score 24.6; DB 1; Length 34806;
Best Local Similarity 87.1%; Pred. No. 3.6;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CCGGTACGGTTCAATCCCATCAGTCACCC 33
Db 32511 CAGGTATGGTTCGAGTCCCATCAGTCACCC 32481

RESULT 15
AX338976      255 bp      DNA      linear      PAT 09-JAN-2002
DEFINITION   Sequence 7 from Patent WO0187936.
ACCESSION   AX338976
VERSION      AX338976.1 GI:18129112
KEYWORDS     Micromonospora halophytica
SOURCE       Micromonospora halophytica
ORGANISM     Micromonospora halophytica
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporineae; Micromonosporaceae; Micromonospora.
REFERENCE    1
AUTHORS      Hosted,T.J. and Horan,A.C.
TITLE        Isolation of Micromonospora carbonacea var africana pmlp1 integrase
and use of integrating function for site-specific integration into
Micromonospora halophytica and Micromonospora carbonacea chromosome
Patent: WO 0187936-A 7 22-NOV-2001;
SCHERING CORPORATION (US)
JOURNAL
FEATURES
source      1..255
            /organism="Micromonospora halophytica"
            /mol_type="unassigned DNA"
            /db_xref="taxon:47864"
ORIGIN
Query Match      70.6%; Score 24; DB 6; Length 255;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GGGTTCAATTCCTCATCAGTCACCC 33
Db 96 GGGTTCAATTCCTCATCAGTCACCC 119

Search completed: August 6, 2004, 11:53:13
Job time : 159.122 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 08:18:49 ; Search time 20.205 Seconds
(without alignments)
7148.662 Million cell updates/sec

Title: US-09-855-340A-3
Perfect score: 34
Sequence: 1 cccggtagcgggttcaattcccatcagtcacccg 34

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_25Jan04 : *
1: Geneseqn1980s : *
2: Geneseqn1990s : *
3: Geneseqn2000s : *
4: Geneseqn2001as : *
5: Geneseqn2001bs : *
6: Geneseqn2002s : *
7: Geneseqn2003as : *
8: Geneseqn2003bs : *
9: Geneseqn2003cs : *
10: Geneseqn2004s : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	34	6	Aad25933 Micromono
2	34	100.0	209	6	Aad25945 Micromono
3	34	100.0	247	6	Aad25936 Micromono
4	34	100.0	4388	5	Aas08694 Micromono
5	33	97.1	260	6	Aad25939 Micromono
6	25	73.5	241	6	Aad25934 Micromono
7	25	73.5	243	6	Aad25935 Micromono
8	25	73.5	315	6	Aad25938 Micromono
9	24	70.6	255	6	Aad25937 Micromono
10	22.4	65.9	110000	2	Continuation (5 of
C 11	20.6	60.6	5546	4	Ab112144 Drosophila
C 12	20.4	60.0	40116	7	Ab226080 Mouse Dna
13	19.6	57.6	348	5	Aaf15421 Thalassio
14	19.6	57.6	348	5	Aaf15526 Thalassio
C 15	19.6	57.6	438	8	Ach26679 Human adu
C 16	19.6	57.6	556	8	Ach28010 Human adu
C 17	19.6	57.6	40681	6	Continuation (7 of
C 18	19.6	57.6	171936	6	Ab56565 Human SUL
C 19	19.4	57.1	532	3	Aaa62850 Human pro
C 20	19.4	57.1	6573	5	Aah4107 Human per
C 21	19.4	57.1	9883	4	Aak79597 Human inn
C 22	19.4	57.1	17601	6	Ab152077 Human per
C 23	19.4	57.1	50196	4	Aak79598 Human inn

24	19.2	56.5	76	6	ABK81114
25	19.2	56.5	76	6	ABK81134
26	19.2	56.5	171	6	ABK79564
27	19.2	56.5	546	4	Aah07734 Human cdn
28	19.2	56.5	744	4	AAI95794
29	19.2	56.5	839	4	Aah84379 Human neu
C 30	19.2	56.5	987	2	AAX13475 Enterococ
C 31	19.2	56.5	987	6	ABS99270 Enterococ
32	19.2	56.5	1433	4	Aah14935 Human cdn
33	19.2	56.5	1723	6	ABL90292 Human pol
34	19.2	56.5	4443	5	AAS86629 DNA encod
35	19.2	56.5	19217	2	AA20518 Polynucle
36	19.2	56.5	41936	8	ADA02693 Mouse Tk2
37	19.2	56.5	41936	9	ADB72431 Mouse Flt
38	19.2	56.5	47115	8	ADA02627 Mouse Flt
39	19.2	56.5	47115	9	ADB72365 Mouse Flt
40	19.2	56.5	134525	2	AAQ04525 Total bas
C 41	19.2	56.5	134525	2	AAQ04525 Total bas
C 42	19	55.9	1581	7	ABT23503 Cellobioh
C 43	19	55.9	1680	3	Aaz65283 Human sec
C 44	19	55.9	1681	7	ADA40018 Human sec
C 45	19	55.9	1681	8	ADB91229 Human sec

ALIGNMENTS

RESULT 1
AAD25933
ID AAD25933 standard; DNA; 34 BP.
XX AC AAD25933;
XX AC AAD25933;
DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attP site DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
KW site-specific integration; hybrid antibiotic; metabolic product;
KW secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX PN WO200187936-A2.
XX PD 22-NOV-2001.
XX PF 15-MAY-2001; 2001WO-US015760.
XX PR 17-MAY-2000; 2000US-0204670P.
XX (SCHE) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX Novel polynucleotides encoding integrase, excisionase and an integrase
XX attachment site isolated from a lysogenic phase pMLP1, useful for
XX transforming an actinomycete.
XX Claim 5; Page 33; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase
XX (int) and excisionase (xis) and an integrase attachment site (attP) which
XX are isolated from pMLP1, a bacteriophage (lysogenic phase) isolated from
XX Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX are useful for transforming an actinomycete with a vector. They are also
XX useful for creating vectors for site-specific integration into host
XX chromosomes. The integrating vectors are used to express actinomycete
XX genes, manipulate secondary metabolic pathways and create new metabolic
XX products such as hybrid antibiotics. The present sequence is pMLP1 attP
XX site DNA from Micromonospora carbonacea

```

XX SQ Sequence 34 BP; 6 A; 14 C; 7 G; 7 T; 0 U; 0 Other;
Query Match 100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGTTCAATTCCTCATCAGTCACCG 34
   |||||
Db 1 CCCCGGTACGGTTCAATTCCTCATCAGTCACCG 34

RESULT 2
AAD25945
ID AAD25945 standard; DNA; 209 BP.
XX AC AAD25945;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attP DNA.
XX KW Integrate; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product;
XX KW secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
FT misc_feature 101..125
FT /*tag= a
FT /*note= "attP region"
FT repeat_region 145..162
FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted repeat 1 (IR1)"
FT repeat_region 167..186
FT /*tag= c
FT /*rpt_type= INVERTED
FT /*note= "Inverted repeat 2 (IR2)"
XX WO200187936-A2.
XX 22-NOV-2001.
XX 15-MAY-2001; 2001WO-US015760.
XX 17-MAY-2000; 2000US-0204670P.
XX (SCHE ) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX Novel polynucleotides encoding integrase, excisionase and an integrase
XX attachment site isolated from a lysogenic phage pMLP1, useful for
XX transforming an actinomycete.
XX Example 3; Fig 4; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase
XX (int) and excisionase (xis) and an integrase attachment site (attP) which
XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX are useful for transforming an actinomycete with a vector. They are also
XX useful for creating vectors for site-specific integration into host
XX chromosomes. The integrating vectors are used to express actinomycete
XX genes, manipulate secondary metabolic pathways and create new metabolic
XX products such as hybrid antibiotics. The present sequence is pMLP1 attP
XX DNA from Micromonospora carbonacea
XX Sequence 209 BP; 49 A; 64 C; 66 G; 30 T; 0 U; 0 Other;

Query Match 100.0%; Score 34; DB 6; Length 209;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGTTCAATTCCTCATCAGTCACCG 34
   |||||
Db 92 CCCCGGTACGGTTCAATTCCTCATCAGTCACCG 125

RESULT 3
AAD25936
ID AAD25936 standard; DNA; 247 BP.
XX AC AAD25936;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attP/attB right juncture DNA.
XX KW Integrate; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; attB/attP DNA; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
FT misc_feature 101..125
FT /*tag= a
FT /*note= "attP/attB region"
FT repeat_region 130..143
FT /*tag= b
FT /*rpt_type= INVERTED
FT /*note= "Inverted repeat 1 (IR1)"
FT repeat_region 148..161
FT /*tag= c
FT /*rpt_type= INVERTED
FT /*note= "Inverted repeat 2 (IR2)"
FT CDS complement(185..247)
FT /*tag= d
FT /*product= "attB peptide"
XX WO200187936-A2.
XX 22-NOV-2001.
XX 15-MAY-2001; 2001WO-US015760.
XX 17-MAY-2000; 2000US-0204670P.
XX (SCHE ) SCHERING CORP.
XX Hosted TJ, Horan AC;
XX WPI; 2002-082983/11.
XX P-PSDB; AAE15909.
XX Novel polynucleotides encoding integrase, excisionase and an integrase
XX attachment site isolated from a lysogenic phage pMLP1, useful for
XX transforming an actinomycete.
XX Claim 23; Fig 4; 34pp; English.
XX The present invention relates to novel polynucleotides encoding integrase
XX (int) and excisionase (xis) and an integrase attachment site (attP) which
XX are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX are useful for transforming an actinomycete with a vector. They are also
XX useful for creating vectors for site-specific integration into host
XX chromosomes. The integrating vectors are used to express actinomycete
XX genes, manipulate secondary metabolic pathways and create new metabolic
XX products such as hybrid antibiotics. The present sequence is pMLP1
XX attP/attB right juncture DNA from Micromonospora carbonacea

```

XX Sequence 247 BP; 38 A; 82 C; 90 G; 37 T; 0 U; 0 Other;
SQ Query Match 100.0%; Score 34; DB 6; Length 247;
Best Local Similarity 100.0%; Pred. No. 2.6e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCCGTACGGGTTCATTCCTCCATCAGTCACCCG 34
DB 92 CCCCGTACGGGTTCATTCCTCCATCAGTCACCCG 125
RESULT 4
ID AAS08694
XX AAS08694 standard; DNA; 4388 BP.
AC AAS08694;
XX 11-SEP-2003 (revised)
DT 26-SEP-2001 (first entry)
XX Micromonospora DNA encoding integrase enzymes.
DE Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;
KW integrase; ds.
XX Micromonospora sp. ATCC 39149.
FH Key Location/Qualifiers
FT CDS 963..1387
FT /tag= a
FT /product= "Integrase #1"
FT CDS 1394..2572
FT /tag= b
FT /product= "Integrase #2"
FT misc_feature 2570..2799
FT /tag= c
FT /note= "AttB/AttP region of integrase action"
FT misc_recomb 27114..2715
FT /tag= d
FT /label= Insertion_juncture
FT /note= "Site of integrase activity"
XX WO200151639-A2.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US001187.
XX 12-JAN-2000; 2000US-0175751P.
XX (SCHE) SCHERING CORP.
XX Hosted TJ, Horan AC, Wang TX;
XX WPI; 2001-442147/47.
XX P-PSDB; AAU04900, AAU04912.
XX New nucleic acid molecules encoding everninomicin pathway gene products,
XX useful for improving yields of everninomicin, to produce new
XX everninomicin and as probes to identify homologous sequences.
XX Claim 26; Fig 7; 109pp; English.
XX The sequence encodes 2 integrases which permit site specific integration
XX of a vector into an actinomycete, especially a Micromonospora genome.
XX The invention relates to nucleic acids and vectors comprising a M.
XX carbonacea everninomicin biosynthetic pathway resistance gene product
XX useful for selecting for a transfected or transformed host cell. An
XX integrative version of the vector is useful for introducing a
XX everninomicin pathway gene (a bottle-neck gene) into an actinomycete of
XX the genus Micromonospora. The DNA encoding the biosynthetic proteins is
XX useful for synthesising novel everninomicin-related compounds, arising

CC from modifications of the DNA sequence designed to change glycosyl and
CC modified orsellinic acid groups contained in everninomicin, for
CC expressing functional or mutant everninomicin biosynthetic enzyme for
CC evaluation, diagnosis and preferably biosynthesis of everninomicin or
CC other secondary metabolic products, improving the yield of everninomicins
CC and to produce novel everninomicins and also as a hybridisation probe to
CC identify homologous sequences. The encoded polypeptides are useful for
CC combinatorial biosynthesis to generate libraries of orthomycins, e.g.
CC everninomicin analogues/homologues and drug discovery. The DNA encoding
CC the integrase allows for increasing a given gene dosage. The integrative
CC vector can be used to permanently integrate copies of a heterologous gene
CC of choice into chromosomes of different hosts and to integrate genes
CC which increase the yield of known products or to generate novel products
CC such as hybrid antibiotics or other novel secondary metabolites. The
CC vector can also be used to integrate antibiotic resistance genes in order
CC to carry out bioconversions with compounds to which the strain is
CC normally sensitive and is thus useful in fermentation processes involving
CC e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS
CC field)
XX Sequence 4388 BP; 714 A; 1468 C; 1506 G; 594 T; 0 U; 6 Other;
SQ Query Match 100.0%; Score 34; DB 5; Length 4388;
Best Local Similarity 100.0%; Pred. No. 4.3e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCCGTACGGGTTCATTCCTCCATCAGTCACCCG 34
DB 2682 CCCCGTACGGGTTCATTCCTCCATCAGTCACCCG 2715
RESULT 5
AAD25939
ID AAD25939 standard; DNA; 260 BP.
XX AAD25939;
XX 07-AUG-2003 (revised)
DT 26-MAR-2002 (first entry)
XX Micromonospora halophytica pMLP1 attP/attB right junction DNA.
DE Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
KW site-specific integration; hybrid antibiotic; metabolic product; attB;
KW secondary metabolic pathway; attB/attP DNA; ds.
XX Micromonospora halophytica.
XX Key Location/Qualifiers
FH misc_feature 101..126
FT /tag= a
FT /note= "attP/attB region"
FT repeat_region 139..150
FT /tag= b
FT /rpt_type= INVERTED
FT /note= "Inverted repeat 1 (IR1)"
FT repeat_region 155..166
FT /tag= c
FT /rpt_type= INVERTED
FT /note= "Inverted repeat 2 (IR2)"
FT CDS complement(196..258)
FT /tag= d
FT /product= "attB peptide".
XX WO200187935-A2.
XX 22-NOV-2001.
XX 15-MAY-2001; 2001WO-US015760.
XX 17-MAY-2000; 2000US-0204670P.
XX (SCHE) SCHERING CORP.

```

XX PI Hosted TJ, Horan AC;
XX DR WPI; 2002-082983/11.
XX DR P-PSDB; AAE15910.
XX PT Novel polynucleotides encoding integrase, excisionase and an integrase
XX PT attachment site isolated from a lysogenic phage pMLP1, useful for
XX PT transforming an actinomycete.
XX PS Claim 23; Fig 5; 34pp; English.
XX CC The present invention relates to novel polynucleotides encoding integrase
XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which
XX CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX CC are useful for transforming an actinomycete with a vector. They are also
XX CC useful for creating vectors for site-specific integration into host
XX CC chromosomes. The integrating vectors are used to express actinomycete
XX CC genes, manipulate secondary metabolic pathways and create new metabolic
XX CC products such as hybrid antibiotics. The present sequence is pMLP1
XX CC attP/attB right junction DNA from Micromonospora halophytica. (Updated on
XX CC 07-AUG-2003 to correct OS field.)
XX SQ Sequence 260 BP; 47 A; 83 C; 90 G; 40 T; 0 U; 0 Other;
Query Match 97.1%; Score 33; DB 6; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCCCGTACGGTTCATTCCTCCATCAGTCACCC 33
Db 92 CCCCGTACGGTTCATTCCTCCATCAGTCACCC 124
RESULT 6
AAD25934
ID AAD25934 standard; DNA; 241 BP.
XX AC AAD25934;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attB DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 44..119
FT /*tag= a
FT /note= "tRNA-His DNA"
FT /*tag= b
FT /note= "attB region"
FT /*tag= c
FT /rpt_type= INVERTED
FT repeat_region 124..137
FT /note= "Inverted repeat 1 (IR1)"
FT /*tag= d
FT repeat_region 142..155
FT /note= "Inverted repeat 2 (IR2)"
FT /*tag= e
FT /product= "attB peptide"
XX CDS
XX WO200187936-A2.
XX 22-NOV-2001.

```

```

XX PF 15-MAY-2001; 2001WO-US015760.
XX PR 17-MAY-2000; 2000US-0204670P.
XX PA (SCHE ) SCHERING CORP.
XX PI Hosted TJ, Horan AC;
XX DR WPI; 2002-082983/11.
XX DR P-PSDB; AAE15909.
XX PT Novel polynucleotides encoding integrase, excisionase and an integrase
XX PT attachment site isolated from a lysogenic phage pMLP1, useful for
XX PT transforming an actinomycete.
XX PS Claim 23; Fig 4; 34pp; English.
XX CC The present invention relates to novel polynucleotides encoding integrase
XX CC (int) and excisionase (xis) and an integrase attachment site (attP) which
XX CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
XX CC Micromonospora carbonacea var. africana. Polynucleotides of the invention
XX CC are useful for transforming an actinomycete with a vector. They are also
XX CC useful for creating vectors for site-specific integration into host
XX CC chromosomes. The integrating vectors are used to express actinomycete
XX CC genes, manipulate secondary metabolic pathways and create new metabolic
XX CC products such as hybrid antibiotics. The present sequence is pMLP1 attB
XX CC DNA from Micromonospora carbonacea
XX SQ Sequence 241 BP; 27 A; 78 C; 92 G; 44 T; 0 U; 0 Other;
Query Match 73.5%; Score 25; DB 6; Length 241;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 10 GGGTTCAATTCCTCCATCAGTCACCCG 34
Db 95 GGGTTCAATTCCTCCATCAGTCACCCG 119
RESULT 7
AAD25935
ID AAD25935 standard; DNA; 243 BP.
XX AC AAD25935;
XX DT 26-MAR-2002 (first entry)
XX DE Micromonospora carbonacea pMLP1 attB/attP left junction DNA.
XX KW Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
XX KW site-specific integration; hybrid antibiotic; metabolic product; attB;
XX KW secondary metabolic pathway; attB/attP DNA; ds.
XX OS Micromonospora carbonacea.
XX FH Key Location/Qualifiers
XX FT misc_feature 62..137
FT /*tag= a
FT /note= "tRNA-His DNA"
FT /*tag= b
FT /note= "attB/attP region"
FT repeat_region 113..137
FT /note= "Inverted repeat 1 (IR1)"
FT /*tag= c
FT repeat_region 179..198
FT /note= "Inverted repeat 2 (IR2)"
FT /*tag= d
XX PN WO200187936-A2.

```

XX	22-NOV-2001.
PD	
XX	
XX	15-MAY-2001; 2001WO-US015760.
PF	
XX	
PR	17-MAY-2000; 2000US-0204670P.
XX	
PA	(SCHE) SCHERING CORP.
XX	
XX	Hosted TJ, Horan AC;
PI	
XX	
DR	WPI; 2002-082983/11.
XX	
PT	Novel polynucleotides encoding integrase, excisionase and an integrase attachment site isolated from a lysogenic phage pMLP1, useful for transforming an actinomycete.
FT	
XX	
XX	Claim 23; Fig 4; 34pp; English.
PS	
CC	The present invention relates to novel polynucleotides encoding integrase (int) and excisionase (xis) and an integrase attachment site (attP) which are isolated from pMLP1, a bacteriophage (lysogenic phage), isolated from Micromonospora carbonacea var. africana. Polynucleotides of the invention are useful for transforming an actinomycete with a vector. They are also useful for creating vectors for site-specific integration into host chromosomes. The integrating vectors are used to express actinomycete genes, manipulate secondary metabolic pathways and create new metabolic products such as hybrid antibiotics. The present sequence is pMLP1 attB/attP left junction DNA from Micromonospora carbonacea
CC	
CC	Sequence 243 BP; 39 A; 74 C; 83 G; 47 T; 0 U; 0 Other;
SQ	
	Query Match 73.5%; Score 25; DB 6; Length 243;
	Best Local Similarity 100.0%; Pred.No. 0.27;
	Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	10 GGTTCAATTCCCATCAGTCACCG 34 113 GGTTCAATTCCCATCAGTCACCG 137
Db	
RESULT 8	
AAD25938	
ID	AAD25938 standard; DNA; 315 BP.
XX	
AC	AAD25938;
XX	
DT	07-AUG-2003 (revised)
DT	26-MAR-2002 (first entry)
XX	
DE	Micromonospora halophytica pMLP1 attB/attP left junction DNA.
XX	
KW	Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
KW	site-specific integration; hybrid antibiotic; metabolic product; attB;
KW	secondary metabolic pathway; attB/attP DNA; ds.
XX	
OS	Micromonospora halophytica.
XX	
FH	Key Location/Qualifiers
FT	misc_feature 45..120
FT	/tag= a
FT	/note= "tRNA-His DNA"
FT	misc_feature 96..120
FT	/tag= b
FT	/note= "attB/attP region"
FT	repeat_region 140..158
FT	/tag= c
FT	/rpt_type= INVERTED
FT	/note= "inverted repeat 1 (IR1)"
FT	repeat_region 162..178
FT	/tag= d
FT	/rpt_type= INVERTED
FT	/note= "inverted repeat 2 (IR2)"
FT	
FT	
XX	
PN	WG200187936-A2.
XX	
XX	22-NOV-2001.
PD	
XX	
XX	15-MAY-2001; 2001WO-US015760.
PF	
XX	
PR	17-MAY-2000; 2000US-0204670P.
XX	
PA	(SCHE) SCHERING CORP.
XX	
XX	Hosted TJ, Horan AC;
PI	
XX	
DR	WPI; 2002-082983/11.
XX	
PT	Novel polynucleotides encoding integrase, excisionase and an integrase attachment site isolated from a lysogenic phage pMLP1, useful for transforming an actinomycete.
FT	
XX	
XX	Claim 23; Fig 5; 34pp; English.
PS	
CC	The present invention relates to novel polynucleotides encoding integrase (int) and excisionase (xis) and an integrase attachment site (attP) which are isolated from pMLP1, a bacteriophage (lysogenic phage), isolated from Micromonospora carbonacea var. africana. Polynucleotides of the invention are useful for transforming an actinomycete with a vector. They are also useful for creating vectors for site-specific integration into host chromosomes. The integrating vectors are used to express actinomycete genes, manipulate secondary metabolic pathways and create new metabolic products such as hybrid antibiotics. The present sequence is pMLP1 attB/attP left junction DNA from Micromonospora halophytica. (Updated on 07-AUG-2003 to correct OS field.)
CC	
CC	Sequence 315 BP; 42 A; 109 C; 101 G; 63 T; 0 U; 0 Other;
SQ	
	Query Match 73.5%; Score 25; DB 6; Length 315;
	Best Local Similarity 100.0%; Pred.No. 0.29;
	Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	10 GGTTCAATTCCCATCAGTCACCG 34 96 GGTTCAATTCCCATCAGTCACCG 120
Db	
RESULT 9	
AAD25937	
ID	AAD25937 standard; DNA; 255 BP.
XX	
AC	AAD25937;
XX	
DT	07-AUG-2003 (revised)
DT	26-MAR-2002 (first entry)
XX	
DE	Micromonospora halophytica pMLP1 attB DNA.
XX	
KW	Integrase; int; excisionase; xis; integrase attachment site; attP; pMLP1;
KW	site-specific integration; hybrid antibiotic; metabolic product; attB;
KW	secondary metabolic pathway; ds.
XX	
OS	Micromonospora halophytica.
XX	
FH	Key Location/Qualifiers
FT	misc_feature 45..121
FT	/tag= a
FT	/note= "tRNA-His DNA"
FT	misc_feature 96..121
FT	/tag= b
FT	/note= "attB region"
FT	repeat_region 134..145
FT	/tag= c
FT	/rpt_type= INVERTED
FT	/note= "Inverted repeat 1 (IR1)"
FT	repeat_region 150..161
FT	

FT FT /*tag= d
FT /rpt_type= INVERTED
FT /notes= "Inverted repeat 2 (IR2)"
CDS complement(191..253)
FT /*tag= e
FT /product= "attB peptide"
FT

XX WO200187936-A2.

XX 22-NOV-2001.

XX 15-MAY-2001; 2001WO-US015760.

XX 17-MAY-2000; 2000US-0204670P.

XX (SCHE) SCHERING CORP.

XX Hosted TJ, Horan AC;

XX WPI; 2002-082983/11.

XX P-PSDB; AAE15910.

XX Novel polynucleotides encoding integrase, excisionase and an integrase
PT attachment site isolated from a lysogenic phage pMLP1, useful for
PT transforming an actinomycete.

XX Claim 23; Fig 5; 34pp; English.

XX The present invention relates to novel polynucleotides encoding integrase
CC (int) and excisionase (xis) and an integrase attachment site (attP) which
CC are isolated from pMLP1, a bacteriophage (lysogenic phage) isolated from
CC Micromonospora carbonacea var. africana. Polynucleotides of the invention
CC are useful for transforming an actinomycete with a vector. They are also
CC useful for creating vectors for site-specific integration into host
CC chromosomes. The integrating vectors are used to express actinomycete
CC genes, manipulate secondary metabolic pathways and create new metabolic
CC products such as hybrid antibiotics. The present sequence is pMLP1 attB
CC DNA from Micromonospora halophytica. (Updated on 07-AUG-2003 to correct
CC OS field.)

XX Sequence 255 BP; 33 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 70.8%; Score 24; DB 6; Length 255;

Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GGGTTCATTCCCATCAGTCACCC 33

Db 96 GGGTTCATTCCCATCAGTCACCC 119

RESULT 10

AAT58840_4

Continuation (5 of 6) of AAT58840 from base 400001 (Mycoplasma genitalium genome.)

WP Sequence split into 6 fragments LOCUS AAT58840 Accession AAT58840

WP Fragment Name Begin End

WP AAT58840_0 1 110000

WP AAT58840_1 100001 210000

WP AAT58840_2 200001 310000

WP AAT58840_3 300001 410000

WP AAT58840_4 400001 510000

WP AAT58840_5 500001 580073

Query Match 65.8%; Score 22.4; DB 2; Length 110000;

Best Local Similarity 95.8%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 GGGTTCATTCCCATCAGTCACCC 33

Db 45128 GGGTTCATTCCCATCAGTCACCC 45151

RESULT 11

ABLI12144/c
ID ABLI12144 standard; cDNA; 5546 BP.

XX ABLI12144;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30914.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-658860/75.

XX P-PSDB; ABB68041.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 30914; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABE57737-
CC ABE72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 5546 BP; 1398 A; 1240 C; 1213 G; 1695 T; 0 U; 0 Other;

Query Match 60.6%; Score 20.6; DB 4; Length 5546;

Best Local Similarity 85.2%; Pred. No. 45;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TACGGTTCATTCCCATCAGTCACCC 33

Db 2511 TCCGAGTTCAAATCCACCATCACC 2485

RESULT 12

ABZ226080/c

ID ABZ226080 standard; DNA; 40116 BP.

XX AC ABZ226080;

XX 24-MAR-2003 (first entry)

DE Mouse DNaseX encoding genomic DNA SEQ ID NO 7.

KW Mouse; DNase X; apoptosis; systemic lupus erythematosus; cystic fibrosis;
KW acquired immune deficiency syndrome; AIDS; cancer; prostatic atrophy;
KW transgenic; dermatological; immunosuppressive; antiinflammatory;
KW anti-HIV; cytostatic; gene; ds.

XX Thalassiosira weissflogii plastid tRNA, SEQ ID NO:168.
DE ssrA gene; tRNA; bacterium; chloroplast; diatom; detection; detection;
KW identification; quantification; characterisation; nucleic acid array;
KW DNA chip; drug design; treatment monitoring; contamination; ss.
XX Thalassiosira weissflogii.
XX WO200070086-A1.
XX 23-NOV-2000.
XX 15-MAY-2000; 2000WO-IB000066.
XX 14-MAY-1999; 99WO-IB000043.
XX (IRBI-) ENTERPRISE IRELAND T/A BIORESEARCH IRELA.
XX (DNA-) UNIV NAI IRELAND GALWAY.
XX Barry TG, Smith TV;
XX WPI; 2001-025025/03.
XX Use of ssrA gene, tRNA, or fragments of them, as target regions in probe
PT assays for detection of prokaryotic or eukaryotic organisms, and for
PT determination of species.
XX Claim 27; Page 83; 22ipp; English.
XX The invention relates to the ssrA gene or tRNA, an RNA transcript of the
CC ssrA gene, or fragments thereof as target regions in a nucleic acid probe
CC assay for the detection and identification of prokaryotic and/or
CC eukaryotic organisms. The invention also relates to 38 novel ssrA
CC sequences and their tRNA transcripts (claimed), and to pan-bacterial,
CC genus- and species-specific ssrA gene/tRNA-directed probes and PCR
CC primers (claimed). tRNA is a stable, high copy number RNA which is found
CC in all bacteria and is also found in chloroplasts and diatoms. It has a
CC dual function both as a tRNA and as an mRNA and is involved in rescuing
CC truncated mRNAs which have lost stop codons. ssrA genes and tRNA
CC sequences can be used as target regions in nucleic acid probe assays for
CC the detection, identification, or quantification of a prokaryotic or
CC eukaryotic organism. cDNA transcripts of tRNA molecules may also be used
CC as probes for in vitro or in situ nucleic acid hybridisation assays. A
CC fragment of the ssrA gene or a tRNA molecule corresponding to a region
CC of high homology from the 5' end or the 3' end of the DNA molecule can be
CC used as a universal target region in a nucleic acid probe assay, while a
CC fragment of the ssrA gene or a tRNA molecule corresponding to a region
CC of low homology can be used as a target region to distinguish between
CC species and as a target region for the generation of genus-specific
CC probes. These regions may also be used as the basis for amplification
CC primer design. The target regions may be used as the basis of an assay
CC for distinguishing between living and dead prokaryotic or eukaryotic
CC organisms, and in a multiple probe format for broad scale detection
CC and/or identification of prokaryotic or eukaryotic organisms. An ssrA
CC gene probe or a tRNA transcript probe can be linked to a microarray gene
CC chip system for the broad scale high throughput detection and
CC identification of prokaryotic or eukaryotic organisms. A fragment of the
CC ssrA gene or the tRNA transcript can be used in an assay to obtain a DNA
CC profile of a prokaryotic organism and distinguish between strains of the
CC same species. The ssrA gene, the tRNA transcript, DNA complementary to
CC an ssrA gene or tRNA, or a fragment thereof can be used to design an
CC agent directed against infectious prokaryotic or eukaryotic organisms for
CC therapeutic purposes, and target regions may be used to monitor the
CC efficacy of drug therapies against infectious agents. Target regions may
CC also be used to monitor the viability and level of probiotic organisms in
CC the gastrointestinal tract. The methods and nucleic acids and
CC compositions of the invention have applications in medicine, and also in
CC industry (e.g., for assessing bacterial contamination of a foodstuff or
CC an environmental sample). Sequences AAF15443-F15547 represent tRNAs, or
CC fragments thereof, from a wide variety of organisms
XX Sequence 348 BP; 124 A; 47 C; 49 G; 0 T; 128 U; 0 Other;

Query Match 57.6%; Score 19.6; DB 5; Length 348;
Best Local Similarity 61.5%; Pred. No. 77;
Matches 16; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 8 ACGGGTTCAATTCCTCAGTCACCC 33
DB 321 AUGGGUCAAUCCCAUCAGUCCAC 346
RESULT 15
ACH26679/C
ID ACH26679 standard; cDNA; 438 BP.
XX AC ACH26679;
XX 13-OCT-2003 (first entry)
XX Human adult ovary cDNA #5059.
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX Homo sapiens.
XX US2003073623-A1.
XX 17-APR-2003.
XX 30-JUL-2001; 2001US-00918995.
XX 30-JUL-2001; 2001US-00918995.
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX Claim 1; SEQ ID NO 13891; 44pp; English.
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH2789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX Sequence 438 BP; 150 A; 73 C; 81 G; 131 T; 0 U; 3 Other;
SQ Query Match 57.6%; Score 19.6; DB 8; Length 438;
Best Local Similarity 84.6%; Pred. No. 80;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 5 GGTACGGGTTCAATTCCTCAGTCA 30
|||
Db 313 GGTTCGGTTCATTCCTCAGTCA 288
|||

Search completed: August 6, 2004, 09:47:34
Job time : 23.205 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:31:14 ; Search time 4.08664 Seconds
(without alignments)
4617.079 Million cell updates/sec

Title: US-09-855-340A-3

Perfect score: 34

Sequence: 1 cccggtaacgggttcaattccatcagtcacccg 34

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.4	65.9	580073	4	US-08-545-528D-1
2	19.6	57.6	640681	4	US-09-790-988-1
3	19.2	56.5	306	4	US-09-134-000C-892
4	19.2	56.5	839	4	US-09-711-164-7
5	18.8	55.3	596	2	US-08-815-175-2
6	18.8	55.3	6305	4	US-09-221-017B-256
7	18.8	55.3	9511	1	US-07-925-695-6
8	18.8	55.3	9511	1	US-07-925-695-7
9	18.6	54.7	1538	4	US-09-644-460-28
10	18.6	54.7	2463	1	US-08-307-444A-10
11	18.6	54.7	2463	1	US-08-587-389-10
12	18.6	54.7	3466	4	US-09-880-484D-1
13	18.6	54.7	3653	4	US-09-023-655-1326
14	18.6	54.7	35100	2	US-08-770-379-17
15	18.6	54.7	35100	3	US-08-757-669A-17
16	18.6	54.7	35100	4	US-09-230-371A-17
17	18.6	54.7	4403765	3	US-09-103-840A-2
18	18.6	54.7	4411529	3	US-09-103-840A-1
19	18.4	54.1	1894	3	US-09-329-350-32
20	18.2	53.5	1281	4	US-09-489-039A-5674
21	18.2	53.5	1356	4	US-09-489-039A-359
22	18.2	53.5	2751	1	US-08-153-848-23
23	18.2	53.5	2751	3	US-09-299-843A-23
24	18.2	53.5	2751	4	US-09-088-337B-23
25	18.2	53.5	2751	5	PCT-US93-11153-23
26	18.2	53.5	5039	4	US-09-386-816C-1
27	18	52.9	472	4	US-09-548-938A-9

C	28	18	52.9	4451	4	US-09-548-938A-1	Sequence 1, Appli
	29	17.8	52.4	213	4	US-09-489-039A-1885	Sequence 1885, Ap
	30	17.8	52.4	213	4	US-09-489-039A-4617	Sequence 4617, Ap
	31	17.8	52.4	213	4	US-09-489-039A-4623	Sequence 4623, Ap
	32	17.8	52.4	213	4	US-09-489-039A-4636	Sequence 4636, Ap
	33	17.8	52.4	213	4	US-09-489-039A-5524	Sequence 5524, Ap
	34	17.8	52.4	213	4	US-09-489-039A-6553	Sequence 6553, Ap
	35	17.8	52.4	240	4	US-09-489-039A-6556	Sequence 6556, Ap
	36	17.8	52.4	252	4	US-09-540-236-802	Sequence 802, App
	37	17.8	52.4	276	4	US-09-252-991A-7369	Sequence 7369, Ap
	38	17.8	52.4	282	4	US-09-328-352-1517	Sequence 1517, Ap
	39	17.8	52.4	354	4	US-09-489-039A-1586	Sequence 1586, Ap
	40	17.8	52.4	354	4	US-09-489-039A-4436	Sequence 4436, Ap
	41	17.8	52.4	354	4	US-09-489-039A-4620	Sequence 4620, Ap
	42	17.8	52.4	354	4	US-09-489-039A-4627	Sequence 4627, Ap
	43	17.8	52.4	354	4	US-09-489-039A-5474	Sequence 5474, Ap
	44	17.8	52.4	354	4	US-09-489-039A-6461	Sequence 6461, Ap
C	45	17.8	52.4	2904	4	US-09-465-355-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-545-528D-1

; Sequence 1, Application US/08545528D

; Patent No. 6537773

; GENERAL INFORMATION:

; APPLICANT: Fraser et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment

; Patent No. 6537773

; TITLE OF INVENTION: Thereof, and Uses Thereof

; FILE REFERENCE: EPI93P1

; CURRENT APPLICATION NUMBER: US/08/545,528D

; CURRENT FILING DATE: 1995-10-19

; PRIOR APPLICATION NUMBER: US 08/488,018

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/473,545

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 1

; LENGTH: 580073

; TYPE: DNA

; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Query Match 65.9%; Score 22.4; DB 4; Length 580073;

Best Local Similarity 95.8%; Pred. No. 2.5;

Matches 23; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GGCTTCAATTCCTCATCAGTCACCC 33

Db 445128 GGCTTCAATTCCTCATCAGTCACCC 445151

RESULT 2

US-09-790-988-1/c

; Sequence 1, Application US/09790988

; Patent No. 6632935

; GENERAL INFORMATION:

; APPLICANT: SHIGENOBU, SHUJI

; APPLICANT: WATANABE, HIDEMI

; APPLICANT: HATTORI, MASAHIRA

; APPLICANT: SAKAKI, YOSHIYUKI

; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS

; FILE REFERENCE: 081356/0159

; CURRENT APPLICATION NUMBER: US/09/790,988

; CURRENT FILING DATE: 2001-02-23

; PRIOR APPLICATION NUMBER: JP2000-107160

; PRIOR FILING DATE: 2000-04-07

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 57.6%; Score 19.6; DB 4; Length 640681;
Best Local Similarity 84.6%; Pred. No. 51;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 ACGGTTCAATCCCATCAGTCAACC 33
DB 625858 ATGGGTCAAGTCCCATGACCAACC 625833

RESULT 3

US-09-134-000C-892/c
; Sequence 892, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent version 3.1
; SEQ ID NO 892
; LENGTH: 306
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-892

Query Match 56.5%; Score 19.2; DB 4; Length 306;
Best Local Similarity 87.5%; Pred. No. 9.8;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 GGGTCAATCCCATCAGTCAACC 33
DB 251 GGGTCAATCCCATGCGTCCGCC 228

RESULT 4

US-09-711-164-7
; Sequence 7, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 839
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(839)
; OTHER INFORMATION: n = A,T,C or G
US-09-711-164-7

Query Match 56.5%; Score 19.2; DB 4; Length 839;
Best Local Similarity 72.7%; Pred. No. 13;

Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 CCGGTACGGTTCAATCCCATCAGTCAACC 34
DB 640 CNGGTCAGGTTCAATACCATCGGTAAATCG 672

RESULT 5

US-08-815-175-2
; Sequence 2, Application US/08815175
; Patent No. 5856139
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL PROLINE-RICH ACIDIC PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA: US/08/815,175
; APPLICATION NUMBER: US/08/815,175
; FILING DATE: Filed Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0225 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PANCTUT02
; CLONE: 2235738
US-08-815-175-2

Query Match 55.3%; Score 18.8; DB 2; Length 596;
Best Local Similarity 69.7%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CCGGTACGGTTCAATCCCATCAGTCAACC 33
DB 497 CCCAGTAGGTTTACGGGCCATNAGTGNCCC 529

RESULT 6

US-09-221-017B-256/c
; Sequence 256, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER

```
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5500
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 256:
SEQUENCE CHARACTERISTICS:
LENGTH: 6305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...6305
US-09-221-017B-256

Query Match 55.3%; Score 18.8; DB 4; Length 6305;
Best Local Similarity 76.7%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CCGTACGGGTTCAATTCCCATCAGTCACCC 33
||||| ||||| ||||| ||||| |||||
Db 2051 CCGTACATATTCGTTTACCATCAGTCGCC 2022

RESULT 7
US-07-925-695-6
Sequence 6, Application US/07925695
Patent No. 5428145
GENERAL INFORMATION:
APPLICANT: OKAMOTO, Hiroaki
APPLICANT: NAKAMURA, Tetsuo
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME.
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
TITLE OF INVENTION: DETECTION SYSTEMS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
STREET: 1850 M Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/925.695
FILING DATE: 19920807
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 287402/91

STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5500
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 256:
SEQUENCE CHARACTERISTICS:
LENGTH: 6305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...6305
US-09-221-017B-256

Query Match 55.3%; Score 18.8; DB 4; Length 6305;
Best Local Similarity 76.7%; Pred. No. 35;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 CCGTACGGGTTCAATTCCCATCAGTCACCC 33
||||| ||||| ||||| ||||| |||||
Db 2051 CCGTACATATTCGTTTACCATCAGTCGCC 2022

RESULT 7
US-07-925-695-6
Sequence 6, Application US/07925695
Patent No. 5428145
GENERAL INFORMATION:
APPLICANT: OKAMOTO, Hiroaki
APPLICANT: NAKAMURA, Tetsuo
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME.
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
TITLE OF INVENTION: DETECTION SYSTEMS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
STREET: 1850 M Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/925.695
FILING DATE: 19920807
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 287402/91

STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5500
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 256:
SEQUENCE CHARACTERISTICS:
LENGTH: 6305 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...6305
US-09-221-017B-256

Query Match 55.3%; Score 18.8; DB 1; Length 9511;
Best Local Similarity 66.7%; Pred. No. 39;
Matches 20; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCGGTACGGGTTCAATTCCCATCAGTCAC 31
||||| ||||| ||||| ||||| |||||
Db 4407 CCGGACGGGACACACUCCCCACGUAAC 4436

RESULT 8
US-07-925-695-7
Sequence 7, Application US/07925695
Patent No. 5428145
GENERAL INFORMATION:
APPLICANT: OKAMOTO, Hiroaki
APPLICANT: NAKAMURA, Tetsuo
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOME.
TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES, ANTIGEN, ANTIBODY AND
TITLE OF INVENTION: DETECTION SYSTEMS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Beveridge, DeGrandi, Weillacher & Young
STREET: 1850 M Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/925.695
FILING DATE: 19920807
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 287402/91
```

```
/ FILING DATE: 09-AUG-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: JP 360441/91
/ FILING DATE: 05-DEC-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weillacher, Robert G.
/ REGISTRATION NUMBER: 20,531
/ REFERENCE/DOCKET NUMBER: 06/87-48009
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 659-2811
/ TELEFAX: (202) 659-1462
/ TELEX: WUI 64470
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9511 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-07-925-695-7

Query Match 55.3%; Score 18.8; DB 1; Length 9511;
Best Local Similarity 76.7%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCGGTACGGGTTCATTCCCATCAGTCAC 31
Db 4407 CCGGTACGGTGACAACTCCCCACAGTAAC 4436

RESULT 9
US-09-644-460-28/c
/ Sequence 28, Application US/09644460
/ Patent No. 6657053
/ GENERAL INFORMATION:
/ APPLICANT: Fisher, Paul B.
/ TITLE OF INVENTION: Reciprocal Subtraction Differential
/ FILE REFERENCE: 34587-G-PCT-USA
/ CURRENT APPLICATION NUMBER: US/09/644,460
/ CURRENT FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: PCT/US99/04323
/ PRIOR FILING DATE: 1999-02-26
/ PRIOR APPLICATION NUMBER: US 09/197,889
/ PRIOR FILING DATE: 1998-11-23
/ PRIOR APPLICATION NUMBER: US 09/185,115
/ PRIOR FILING DATE: 1998-11-03
/ PRIOR APPLICATION NUMBER: US 09/032,684
/ PRIOR FILING DATE: 1998-02-27
/ NUMBER OF SEQ ID NOS: 42
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 1538
/ TYPE: DNA
/ ORGANISM: homo sapiens
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: 652, 1523
/ OTHER INFORMATION: c, t, a or g
/ NAME/KEY: misc.feature
/ LOCATION: (1) (1538)
/ OTHER INFORMATION: n = A,T,C or G
US-09-644-460-28

Query Match 54.7%; Score 18.6; DB 4; Length 1538;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCGCGTACGGGTTCATTCCCATCAGTCACCC 33
Db 1240 CCGAGTACAGTGCCAGTGCCCTCAGACGCC 1208

RESULT 10
```

```
US-08-307-444A-10
/ Sequence 10, Application US/08307444A
/ Patent No. 5516659
/ GENERAL INFORMATION:
/ APPLICANT: NII, ATSUSHI
/ APPLICANT: MORISHITA, HIDEAKI
/ APPLICANT: UEMURA, AKIO
/ APPLICANT: MOCHIDA, EI
/ TITLE OF INVENTION: ANTICOAGULANT POLYPEPTIDES
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: OLIFF & BERRIDGE
/ STREET: P.O. BOX 19928
/ CITY: ALEXANDRIA
/ STATE: VA
/ COUNTRY: USA
/ ZIP: 22320
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/307,444A
/ FILING DATE: 19-SEP-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/835,436
/ FILING DATE: 26-FEB-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: OLIFF, JAMES A.
/ REGISTRATION NUMBER: 27,075
/ REFERENCE/DOCKET NUMBER: JAO 27706
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 836-8400
/ TELEFAX: (703) 836-2787
/ TELEX: 90-1799 PTO ALEX
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2463 bases
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
US-08-307-444A-10

Query Match 54.7%; Score 18.6; DB 1; Length 2463;
Best Local Similarity 72.7%; Pred. No. 33;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCGCGTACGGGTTCATTCCCATCAGTCACCC 33
Db 2338 CCGAGAACTGGGCCAACTCAGTCAGTCACCC 2370

RESULT 11
US-08-587-389-10
/ Sequence 10, Application US/08587389
/ Patent No. 5695964
/ GENERAL INFORMATION:
/ APPLICANT: NII, ATSUSHI
/ APPLICANT: MORISHITA, HIDEAKI
/ APPLICANT: UEMURA, AKIO
/ APPLICANT: MOCHIDA, EI
/ TITLE OF INVENTION: TRUNCATED THROMBOMODULIN, RECOMBINANT
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: OLIFF & BERRIDGE
/ STREET: P.O. BOX 19928
/ CITY: ALEXANDRIA
/ (AS AMENDED)
```

STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,389
FILING DATE: 17-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,444
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: OLIFF, JAMES A.
REGISTRATION NUMBER: 27,075
REFERENCE/DOCKET NUMBER: JAO 27706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6400
TELEFAX: (703) 836-2787
TELEX: 90-1799 PTO ALEX
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2463 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-587-389-10

Query Match	54.7%;	Score 18.6;	DB 1;	Length 2463;
Best Local Similarity	72.7%;	Pred. No. 33;		
Matches	24;	Conservative	0;	Mismatches
			9;	Indels
			0;	Gaps
			0;	

```

Qy      1  CCCCGTACGGTTCAAATCCCATCACTCAACC 33
      |||  |||  |||  |||  |||  |||  |||  |||
Db      2338  CCAGGAACGGGCCCAACTCACTGAGTCAACC 2370

RESULT 12
US 09-880-484D-1
; Sequence 1, Application US/09880484D
; Patent No. 6632791
; GENERAL INFORMATION:
; APPLICANT: Light, David
; APPLICANT: Nagashima, Mariko
; APPLICANT: Morser, Michael J
; TITLE OF INVENTION: Thrombomodulin Analogs for Pharmaceutical Use
; FILE REFERENCE: 51863AUSM1
; CURRENT APPLICATION NUMBER: US/09/880,484D
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/213,678
; PRIOR FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3466

```

```

? LENGTH: 3466
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (151)..(1875)
? OTHER INFORMATION:
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: (205)..()
? OTHER INFORMATION:
? US-09-880-494D-1

```

CHIEF, CIVIL RIGHTS DIVISION
U.S. DEPARTMENT OF JUSTICE
WASHINGTON, D.C. 20530

```

Query Match      54.7%; Score 18.6; DB 4; Length 3466;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCCCGGTACGGGTTCAATTCCCATCAGTCACCC 33
Db 2450 CCCAGAACTGGGCCCAACTCACTGAGTCACCC 2482

```

RESULT 13
US-09-023-655-1326
Sequence 1326, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

1 FILING DATE:
 2 CLASSIFICATION:
 3 ATTORNEY/AGENT INFORMATION:
 4 NAME: Zeller, Karen J.
 5 REGISTRATION NUMBER: 37,071
 6 REFERENCE/DOCKET NUMBER: PA-0001 US
 7 TELECOMMUNICATION INFORMATION:
 8 TELEPHONE: (650) 855-0555
 9 TELEFAX: (650) 848-4166
 10 INFORMATION FOR SEQ ID NO: 1326:
 11 SEQUENCE CHARACTERISTICS:
 12 LENGTH: 3653 base pairs
 13 TYPE: nucleic acid
 14 STRANDEDNESS: single
 15 TOPOLOGY: linear
 16 IMMEDIATE SOURCE:
 17 CLONE: GENBANK
 18 U9-09-023-655-1326
 19 US-09-023-655-1326

Query Match	54.7%	Score 18.6;	DB 4;	Length 3653;
Best Local Similarity	72.7%	Pred. No. 37;		
Matches 24: Conservative	0;	Mismatches 9;	Indels 0;	Gaps 0;

1 CCCCGGTACGGGTTCATTCCCATCAGTCACCC 33
2446 CCCAGAACTGGGCCAACTCACCTGAGTCACCC 2478

RESULT 14
US-08-770-379-17/c
; Sequence 17, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan

APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-17

Query Match 54.7%; Score 18.6; DB 2; Length 35100;
Best Local Similarity 72.7%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 CCCCGTACGGGTCAATTCCTCATCAGTCACCC 33
Db 7143 CCCAGGTACGGGTGAGATACCCCTCAGCTTCCC 7111

RESULT 15
US-08-757-669A-17/c
Sequence 17, Application US/08/757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-17
Query Match 54.7%; Score 18.6; DB 3; Length 35100;
Best Local Similarity 72.7%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 CCCCGTACGGGTCAATTCCTCATCAGTCACCC 33
Db 7143 CCCAGGTACGGGTGAGATACCCCTCAGCTTCCC 7111
Search completed: August 6, 2004, 13:42:37
Job time : 10.0866 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2004, 09:22:34 ; Search time 131.685 Seconds
(without alignments)
7710.163 Million cell updates/sec

Title: US-09-855-340A-3
Perfect score: 34
Sequence: 1 cccggtagcgggttcattccatccatcagtcacccg 34

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Lasting first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	22.4	65.9	689	28	AZ933374 BJ_Ba000
C 2	21.8	64.1	663	10	BF165844 601774875
C 3	21.8	64.1	913	13	BQ889122 AGENCOURT
C 4	21.8	64.1	914	10	BF540384 602050148

C 5	21.4	62.9	496	14	CD221929
C 6	21.4	62.9	617	28	BZ253396
C 7	21.4	62.9	619	14	CD228161
C 8	21.2	62.4	512	28	AZ663671
C 9	21.2	62.4	647	28	BH698487
C 10	21.2	62.4	649	28	AZ663662
C 11	20.8	61.2	369	28	AZ658076
C 12	20.8	61.2	425	29	CG962621
C 13	20.8	61.2	459	28	BZ755997
C 14	20.8	61.2	481	28	BZ866579
C 15	20.8	61.2	503	12	BI128480
C 16	20.8	61.2	607	28	BZ440539
C 17	20.8	61.2	772	10	BF166788
C 18	20.8	61.2	772	28	CC335988
C 19	20.8	61.2	797	29	CG963642
C 20	20.8	61.2	842	29	CG826738
C 21	20.8	61.2	846	29	CG932119
C 22	20.8	61.2	881	28	BZ405594
C 23	20.8	61.2	882	12	BI736662
C 24	20.8	61.2	896	29	CG945888
C 25	20.8	61.2	931	29	CG953021
C 26	20.8	61.2	1161	12	BG342311
C 27	20.6	60.6	459	28	AZ407486
C 28	20.6	60.6	499	29	TA134G07Q
C 29	20.6	60.6	589	28	AQ272293
C 30	20.6	60.6	697	28	BH105996
C 31	20.4	60.0	239	9	AV345146
C 32	20.4	60.0	274	9	AV338955
C 33	20.4	60.0	478	10	BE554144
C 34	20.4	60.0	527	28	BH290990
C 35	20.4	60.0	640	10	BE634824
C 36	20.4	60.0	654	10	BB484517
C 37	20.4	60.0	669	14	CB841703
C 38	20.4	60.0	779	13	BU562772
C 39	20.4	60.0	856	10	BB036725
C 40	20.4	60.0	985	9	AV327597
C 41	20.4	60.0	1890	11	AK031860
C 42	20.4	60.0	2893	11	AK019872
C 43	20.4	60.0	3747	11	AK037906
C 44	20.2	59.4	247	12	BI003984
C 45	20.2	59.4	338	9	AA343504

ALIGNMENTS

RESULT 1
AZ933374/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

AZ933374 689 bp DNA linear GSS 24-APR-2001
BJ_Ba0001H19r B. japonicum BAC library Bradyrhizobium japonicum
genomic, genomic survey sequence.
AZ933374
GSS.
Bradyrhizobium japonicum
Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.
1 (bases 1 to 689)
Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Golcochea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
Genome Res. 11 (8), 1434-1440 (2001)
21376150
11483585
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4288
Fax: 864 656 7298
Email: rwing@clemson.edu


```

REFERENCE
  1 (bases 1 to 914)
  NIH-MGC http://mgs.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  JOURNAL
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: cgaabs@mail.nih.gov
    Tissue Procurement: Jeffrey E. Green, M.D.
    CDNA Library Preparation: Life Technologies, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: LL9515 row: 0 column: 12
    High quality sequence stop: 601.
    Location/Qualifiers
      1..914
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone="IMAGE:4189667"
        /lab_host="DH10B (TI phage-resistant)"
        /clone_lib="NCI CGAP SG2"
        /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
        NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
        dt. Average insert size 1.3 kb. Constructed by Life
        Technologies. Note: this is a NCI CGAP Library."

ORIGIN
  Query Match      64.1%; Score 21.8; DB 10; Length 914;
  Best Local Similarity 78.8%; Pred. NO. 2.4e+02;
  Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCGGTACGGTTCATTCATCCATCAGTCACCC 33
    |||||
Db 602 CCGGTACGGTTCATTCATCCATCAGTCACCC 634

RESULT 5
CD221929/c
LOCUS
DEFINITION
  CCC1_1_E04_b1_A007 Callus culture/cell suspension Sorghum bicolor
  cDNA clone CCC1_1_E04_A007 3', mRNA sequence.
CD221929
VERSION
CD221929.1 GI:30965363
EST.
Sorghum bicolor (sorghum)
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 496)
Klein,R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K.,
Eastman,A. and Pratt,L.H.
An EST database from Sorghum: callus culture and cell suspension
Unpublished (2003)
Other ESTs: CCC1_1_E04.G1_A007
Contact: Cordonnier-Pratt WM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to

```

```

exclude polyA.
Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)
POLYA=Yes.
FEATURES
  source
    1..496
      /organism="Sorghum bicolor"
      /mol_type="mRNA"
      /cultivar="RTX430"
      /db_xref="taxon:4558"
      /clones="CCC1_1_E04_A007"
      /lab_host="DH10B-TI phage-resistant E. coli"
      /clone_lib="Callus culture/cell suspension"
      /note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
      library was prepared from a mixture of polyA+ RNA from
      callus culture tissue and cells in suspension culture.
      Double-stranded cDNA was cloned unidirectionally into
      different DraIII sites of the pME18S-FL3 vector (5-prime
      DraIII site is CACTGTGTG, 3-prime DraIII site is
      CACCATGTG). XhoI excises the cDNA insert."

ORIGIN
  Query Match      62.9%; Score 21.4; DB 14; Length 496;
  Best Local Similarity 80.6%; Pred. No. 3.2e+02;
  Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CCGGTACGGTTCATTCATCCATCAGTCACCC 33
    |||||
Db 255 CCGGTACGGTTCATTCATCCATCAGTCACCC 225

RESULT 6
BZ253996
LOCUS
DEFINITION
  BZ230-307C21-TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
  CH230-307C21, genomic survey sequence.
BZ253996
ACCESSION
VERSION
BZ253996.1 GI:23957934
KEYWORDS
SOURCE
  Rattus norvegicus (Norway rat)
  Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 617)
  Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
  Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
  Riggs,F. de Jong,P. and Fraser,C.M.
  Rat BAC End Sequences from Library CHORI-230 MboI segment
  Unpublished (1999)
  Other GSs: CH230-307C21.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the rat BAC library CHORI-230
  (http://www.chori.org/bacpac/rat230.htm). For BAC library
  availability, please contact Pieter de Jong (pdejong@mail.cho.org).
  Clones may be purchased from BACPAC Resources
  (http://www.chori.org/bacpac/orering_information.htm). BAC end
  page: http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html
  Plate: 307 row: C column: 21
  Seq primer: T7
  Class: BAC ends.
  Location/Qualifiers
    1..617
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /strain="BN/SsNHsd/MCW"
      /db_xref="taxon:10116"
      /clone="CH230-307C21"

```

/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Ract (5N/5Nrd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 62.9%; Score 21.4; DB 28; Length 619;
Best Local Similarity 80.6%; Pred. No. 3.3e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCCCGTACGGTTCATTCATCCCATCAGTCAC 31
|||||
Db 261 CCCAGGTTCCGGTTCATTCCTAACACCCAC 291

RESULT 7
CD228161/c
LOCUS
DEFINITION
CC1_6_C06_b1_A007 Callus culture/cell suspension Sorghum bicolor
cDNA clone CC1_6_C06_A007 3', mRNA sequence.

ACCSSSTON
VERSION
CD228161.1 GI:30971595
KEYWORDS
EST.

Sorghum bicolor (sorghum)

SOURCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Sorghum.

REFERENCE

AUTHORS
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S.,
Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Rathore, K.,
Eastman, A. and Pratt, L.H.

TITLE
An EST database from Sorghum: callus culture and cell suspension

JOURNAL
Unpublished (2003)

COMMENT
Other ESTs: CC1_6_C06_G1_A007

Contact: Cordonnier-Pratt, M.M.
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210

Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo Institute of Medical
Science; plant material and RNA prepared at Texas A & M University;
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.

Seq primer: Sug3 (CGACCTGCAGCTCGAGCACA)

FEATURES

source

Location/Qualifiers

1. 619
/organism="Sorghum bicolor"
/mol_type="mRNA"
/culturivar="RTX430"
/db_xref="taxon:4558"

/clone="CC1_6_C06_A007"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="Callus culture/cell suspension"
/note="Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The
library was prepared from a mixture of polyA+ RNA from
callus culture tissue and cells in suspension culture.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACGTGTGTG, 3-prime DraIII site is
CACCATGTGT). XhoI excises the cDNA insert."

ORIGIN

Query Match 62.9%; Score 21.4; DB 14; Length 619;

Best Local Similarity 80.6%; Pred. No. 3.3e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3 CCCGTACGGTTCATTCATCCCATCAGTCACCC 33
|||||
Db 382 CCCACAGGTTCCGATTCATCCCATCCTTCACCC 352

RESULT 8

AZ663671

LOCUS

DEFINITION

1M0543K22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0543K22 F, genomic survey sequence.

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 512)

Dunn, B., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
Niederhausen, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112 USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0543 row: K column: 22

Seq primer: CGTGTAAACGACGGCCAGT

Class: Plasmid ends

High quality sequence stop: 512.

Location/Qualifiers

1. 512

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0543K22"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 [gi|4732114|gb|AF129072.1], a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adaptor mouse DNA was annealed to

adaptor vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

ORIGIN

Query Match 62.4%; Score 21.2; DB 28; Length 512;

Best Local Similarity 88.5%; Pred. No. 3.9e+02; Mismatches 3; Indels 0; Gaps 0;
 Matches 23; Conservative 0;

QY 6 GTACGGTTCATTCCTCCATCAGTCAC 31
 |||||
 Db 299 GAACGTGTTCAATTCCTCCAGCAGTCAC 324

RESULT 9
 BH698487
 LOCUS
 DEFINITION
 BOHXM56TR BO_2.3_KB Brassica oleracea genomic clone BOHXM56,
 genomic survey sequence.
 ACCESSION
 BH698487
 VERSION
 BH698487.1 GI:18772681
 KEYWORDS
 GSS.
 SOURCE
 Brassica oleracea
 ORGANISM
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 647)
 REFERENCE
 Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 JOURNAL
 Other GSSs: BOHXM56TF
 COMMENT
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
 source
 1..647
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOHXM56"
 /clone_lib="BO_2.3_KB"
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pHO51 using BstXI linkers"

ORIGIN
 Query Match 62.4%; Score 21.2; DB 28; Length 647;
 Best Local Similarity 88.5%; Pred. No. 4e+02; Mismatches 3; Indels 0; Gaps 0;
 Matches 23; Conservative 0;

QY 9 CGGGTTCATTCCTCCATCAGTCACCG 34
 |||||
 Db 618 CGGGTTCATTCCTCCGTCGTCGCGCG 643

RESULT 10
 AZ663662
 LOCUS
 DEFINITION
 1M0543124F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0543124 F, genomic survey sequence.
 ACCESSION
 AZ663662
 VERSION
 AZ663662.1 GI:11800808
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 649)
 REFERENCE
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,R., Rose,R., Stokes,R., Tingey,A., von
 Niederhausern,A. and Wright,D.,Weiss,R.

TITLE
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL
 Unpublished (2000)
 COMMENT
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0543 row: I column: 24
 Seq primer: CGTGTAAACGACGCGCAGT
 Class: plasmid ends
 High quality sequence stop: 649.

FEATURES
 source
 1..649
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0543124"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: pWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN
 Query Match 62.4%; Score 21.2; DB 28; Length 649;
 Best Local Similarity 88.5%; Pred. No. 4e+02; Mismatches 3; Indels 0; Gaps 0;
 Matches 23; Conservative 0;

QY 6 GTACGGTTCATTCCTCCATCAGTCAC 31
 |||||
 Db 300 GAACGTGTTCAATTCCTCCAGCAGTCAC 325

RESULT 11
 AZ658076
 LOCUS
 DEFINITION
 1M0534H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0534H08 R, genomic survey sequence.
 ACCESSION
 AZ658076
 VERSION
 AZ658076.1 GI:11795222
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 369)
 REFERENCE
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,R., Rose,R., Stokes,R., Tingey,A., von
 Niederhausern,A. and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: rdunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0534 row: H column: 08
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 369.

FEATURES Location/Qualifiers
source 1..369
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0534H08"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G14732114|GB|A129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 61.2%; Score 20.8; DB 28; Length 369;
Best Local Similarity 91.7%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 GGTTCAATTCCTCATGTCACCC 33
Db 212 GGTTCAATTCCTCATGTCACAC 235

RESULT 12
CG942621/c
LOCUS MBEFX20TR mth2 Medicago truncatula genomic clone 47D16, genomic survey sequence.
DEFINITION CG942621
ACCESSION CG942621
VERSION CG942621.1 GI:39885464
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 425)
TOWN, C.D., SHERRY, J., KOO, H. and FELDBLYUM, T.F.
AUTHORS Sequencing of BAC ends from Medicago truncatula

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: rdunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0534 row: H column: 08
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 369.

FEATURES Location/Qualifiers
source 1..369
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0534H08"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G14732114|GB|A129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 61.2%; Score 20.8; DB 28; Length 369;
Best Local Similarity 91.7%; Pred. No. 5.4e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 10 GGTTCAATTCCTCATGTCACCC 33
Db 212 GGTTCAATTCCTCATGTCACAC 235

RESULT 12
CG942621/c
LOCUS MBEFX20TR mth2 Medicago truncatula genomic clone 47D16, genomic survey sequence.
DEFINITION CG942621
ACCESSION CG942621
VERSION CG942621.1 GI:39885464
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 425)
TOWN, C.D., SHERRY, J., KOO, H. and FELDBLYUM, T.F.
AUTHORS Sequencing of BAC ends from Medicago truncatula

JOURNAL COMMENT

Unpublished (2003)
Other GSSs: MBEFX20TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3323
Fax: 301-838-0208
Email: cdtown@tigr.org
Seq primer: CAGGAAACAGCTATGACC
Class: BAC ends.

FEATURES source

Location/Qualifiers
1..425
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="47D16"
/clone_lib="mth2"
/note="Vector: pBelOBAC11; Site 1: HindIII; Site 2: HindIII; Cook, D.R. and Kim, D.J., unpublished"

ORIGIN

Query Match 61.2%; Score 20.8; DB 29; Length 425;
Best Local Similarity 78.1%; Pred. No. 5.5e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCCGGTACGGTTCAATTCCTCATGTCACCC 33
Db 158 CCATGCGCGGTTCAATTCCTCATGTCGCC 127

RESULT 13

BZ755997
LOCUS PUBBL38TD ZM_0.6_1.0 KB Zea mays genomic clone ZMMBTa281G04, genomic survey sequence.
DEFINITION BZ755997
ACCESSION BZ755997.1 GI:28908378
VERSION BZ755997
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 459)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw

TITLE

Unpublished (2003)
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: Tg
Class: sheared ends.

FEATURES source

Location/Qualifiers
1..459
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTa281G04"
/clone_lib="ZM 0.6_1.0 KB"
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high COT selected genomic DNA library"

ORIGIN

Query Match 61.2%; Score 20.8; DB 28; Length 459;
Best Local Similarity 78.1%; Pred. No. 5.6e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 2 CCGGTACGGGTTCATTCCTCCATCAGTCACCC 33
    |||
Db 145 CCATGCGGTGTTCATTCCTCCATCATTCAACC 176
    |||

RESULT 14
BZ866579/c
LOCUS BZ866579 481 bp DNA linear GSS 18-MAR-2003
DEFINITION CH240_287E17.TV CHORI-240 Bos taurus genomic clone CH240_287E17,
           genomic survey sequence.
ACCESSION BZ866579
VERSION BZ866579.1 GI:29093984
KEYWORDS GSS.
SOURCE Bos taurus (cow)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
    Bovidae; Bovinae; Bos.
  REFERENCE
    1 (bases 1 to 481)
    Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
    Shwartsbeyn, A., Gebregorgis, E., Chen, D., Riggs, F., de Jong, P.,
    Crawford, A.M. and McEwan, J.C.
    Bovine BAC End Sequences from Library CHORI-240
    Unpublished (2003)
  TITLE Contact: Shaving Zhao
  JOURNAL Department of Eukaryotic Genomics
  COMMENT The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0200
           Fax: 301 838 0208
           Email: szhao@tigr.org
           Clones are derived from the bovine BAC library CHORI-240
           (http://www.chori.org/bacpac/bovine240.htm). For BAC library
           availability, please contact Pieter de Jong (pdejong@mail.cho.org).
           Clones may be purchased from BACPAC Resources
           (http://www.chori.org/bacpac/ordering\_information.htm). This work
           was undertaken as part of the International Bovine BAC Mapping
           Consortium (IBBMC) by AgResearch Ltd., New Zealand and the
           Institute of Genomic Research (TIGR), USA.
           Plate: 287 row: E column: 17
           Seq primer: 17
           Class: BAC ends.
  FEATURES
    Location/Qualifiers
      1..481
        /organism="Bos taurus"
        /mol_type="genomic DNA"
        /strain="breed: Hereford"
        /db_xref="taxon:9913"
        /clone="CH240_287E17"
        /sex="Male"
        /cell_type="Blood"
        /clone_lib="CHORI-240"
        /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
        Hereford Bull L1 Domino 99375; CHORI-240 Bovine BAC
        library (Male) produced by Pieter de Jong"

ORIGIN
  Query Match 61.2%; Score 20.8; DB 28; Length 481;
  Best Local Similarity 78.1%; Pred. No. 5.6e+02;
  Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCGCGGTACGGGTTCATTCCTCCATCAGTCACC 32
    |||
Db 114 CCGCGGTACACCTTGACTTCCCATGAATCAACC 83
    |||

RESULT 15
B1128480/c
LOCUS B1128480 503 bp mRNA linear EST 31-DEC-2001
DEFINITION G076F43Y Populus cambium cDNA library Populus tremula x Populus
           tremuloides cdna, mRNA sequence.
ACCESSION B1128480

```

```

VERSION B1128480.1 GI:18012435
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
  ORGANISM
    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids I; Malpighiales; Salicaceae; Populus.
  REFERENCE
    1 (bases 1 to 503)
    Hertzberg, M., Aspeborg, H., Erlandsson, R., Bjorkbacka, H.,
    Hiltunen, T., Karlsson, J., Teeri, T., Gustafsson, P., Bahlér, R.,
    Jansson, S., Nilsson, C., Sundberg, B., Nilsson, P., Uhlen, M.,
    Sandberg, G. and Lundberg, J.
    Gene expression in Populus
    Unpublished (2001)
  TITLE Contact: Erlandsson R
  JOURNAL Department of Biotechnology
  COMMENT Royal Institute of Technology
           Teknikringen 30, Stockholm S-10044, Sweden
           Tel: 46 8 790 8287
           Fax: 46 8 245452
           Email: rikerl@biochem.kth.se.
  FEATURES
    source
      1..503
        /organism="Populus tremula x Populus tremuloides"
        /mol_type="mRNA"
        /db_xref="taxon:47664"
        /clone_lib="Populus cambium cDNA library"
        /note="Organ: cambium"

ORIGIN
  Query Match 61.2%; Score 20.8; DB 12; Length 503;
  Best Local Similarity 78.1%; Pred. No. 5.6e+02;
  Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCGGTACGGGTTCATTCCTCCATCAGTCACCC 33
    |||
Db 426 CCATGCGGTTCATTCCTCCATCAGTCACCC 395
    |||

Search completed: August 6, 2004, 13:39:06
Job time : 135.685 secs

```

This Page Blank (uspto)